JOURNAL OF RESEARCH ON EDUCATIONAL EFFECTIVENESS
2022, VOL. 15, NO. 1, 130–150
https://doi.org/10.1080/19345747.2021.1953200

Routledge
Taylor & Francis Group

METHODOLOGICAL STUDIES

Optimal Sample Allocation for Three-Level Multisite Cluster-Randomized Trials

Zuchao Shen[a] and Benjamin Kelcey[b]

[a]College of Education, University of Florida, Gainesville, Florida, USA; [b]Quantitative Research Methodologies Program, School of Education, University of Cincinnati, Cincinnati, Ohio, USA

ABSTRACT

Optimal sampling frameworks attempt to identify the most efficient sampling plans to achieve an adequate statistical power. Although such calculations are theoretical in nature, they are critical to the judicious and wise use of funding because they serve as important starting points that guide practical discussions around sampling tradeoffs and requirements. Conventional optimal sampling frameworks, however, often identify sub-optimal designs because they typically presume the costs of sampling units are equal across treatment conditions. In this study, we develop a more flexible framework that allows costs to differ by treatment conditions and derive the optimal sample size formulas for three-level multisite cluster-randomized trials. We find that the proposed optimal sampling schemes are driven by the differences in costs between treatment conditions, cross-level sampling cost ratios and cross-level variance decomposition ratios. We illustrate the utility of the proposed framework by comparing it to a conventional framework and find that the proposed framework frequently identifies more efficient designs. The proposed optimal sampling framework has been implemented in the *R* package *odr*.

ARTICLE HISTORY
Received 21 June 2020
Revised 21 April 2021
Accepted 29 April 2021

KEYWORDS
Optimal sample allocation; optimal design; design efficiency; power analysis; multisite cluster-randomized trials

Multisite cluster-randomized trials blend multisite and hierarchical designs by assigning treatment conditions at random to intermediate clusters (e.g., classrooms) of individuals (e.g., students) within each site (e.g., school). Two recent studies (Spybrook et al., 2016; Spybrook & Raudenbush, 2009) found that, among all types of multilevel experiments funded by the Institute of Education Sciences in the U.S. Department of Education, three-level multisite cluster-randomized trials were among the most frequently used designs because they have many desirable practical and statistical properties. For example, these designs often align well with the typical hierarchical structure of education (e.g., students nested whinin classes nested within schools) and the nature of many classroom-based treatments. Three-level multisite designs essentially replicate a small cluster-randomized study across many sites, they create opportunities to learn about treatment effect variation. Although there are multiple practical and statistical advantages to these designs, these advantages are often tempered by the complexity of the design and the difficulty in identifying efficient sample allocations across the levels when resources are limited (e.g., Raudenbush & Liu, 2000).

CONTACT Zuchao Shen zuchao.shen@gmail.com College of Education, University of Florida, Gainesville, FL 32611, USA.

When planning multisite cluster-randomized trials, investigators must consider both the power of a particular sampling plan and the relative efficiency with which it uses resources (Hedges & Borenstein, 2014; Shen & Kelcey, 2020b, 2020c). The intersection of these criteria has typically been addressed through optimal sampling frameworks that seek to identify sample allocations that maximize power or efficiency while maintaining a fixed set of costs (Liu, 2003; Raudenbush, 1997). Conventional optimal sampling frameworks have covered a variety of multilevel designs, including cluster-randomized trials (Hedges & Borenstein, 2014; Konstantopoulos, 2009, 2011; Raudenbush, 1997), multisite-randomized trials (Raudenbush & Liu, 2000), and multisite cluster-randomized designs (Hedges & Borenstein, 2014). Implicit in these conventional frameworks are two important assumptions that may constrict their utility: (a) sampling costs for units (e.g., students, classes or schools) are equal across treatment conditions and (b) assignment to treatment condition is necessarily optimized under a balanced design (i.e., 50% to treatment, 50% to control; Hedges & Borenstein, 2014; Konstantopoulos, 2009, 2011; Raudenbush, 1997; Raudenbush & Liu, 2000).

These assumptions may not be reasonable in practice under many settings. For instance, holding constant the unit sampling cost across conditions precludes the possibility that the treatment condition incurs extra costs associated with its implementation and delivery. This is not always the case. For example, for Springer et al. (2011), when studying the impact of teacher performance bonuses, each additional teacher placed into the treatment condition incurred a much higher cost (up to US$15,000 a year) than each additional control teacher. Similarly, class size studies and teacher aid studies have documented the differential costs associated with small classes, regular classes, and regular classes paired with teaching aids (Mosteller, 1995). More generally, such differential costs are not unique to education. Differential cost structures have been noted in many areas of health, psychology, and other social sciences because of, for example, the cost of training and implementing interventions (e.g., Greenleaf et al., 2011; Hiscock et al., 2008; Jacob et al., 2015; Jayanthi et al., 2017; Jennings et al., 2017).

Probing the assumption of equal unit costs across conditions also raises downstream questions about the universal efficacy of balanced assignment of units to treatment conditions. Conventional optimal sampling frameworks have regularly concluded that the balanced assignment of units to treatment conditions (i.e., 50% treatment, 50% control) will always produce the most efficient design (e.g., Raudenbush, 1997; see also Hedges & Borenstein, 2014; Konstantopoulos, 2009, 2011). However, as we shall see, the optimality of balanced assignment to treatment conditions is heavily contingent upon the equality of costs assumption. That is, once we have relaxed the equal unit cost across treatment conditions assumption, balanced assignment does not necessarily provide the most efficient design.

To address these limitations, recent literature has introduced more flexible optimal sampling frameworks for two- and three-level cluster-randomized trials (Shen & Kelcey, 2020c) and two-level multisite experiments (Shen & Kelcey, 2020b) that allow sampling costs to vary across both levels of hierarchy and treatment conditions. This literature has demonstrated that when costs differ across treatment conditions, there are more efficient designs than those derived from the conventional optimal design frameworks. Indeed, the conventional frameworks that confine their scope to equal cost settings (e.g., Konstantopoulos, 2009, 2011; Raudenbush, 1997; Raudenbush & Liu, 2000) are special and constrained cases of the more flexible frameworks.

We extend this more flexible optimal sampling framework to the widely used three-level multisite cluster-randomized design. Section I develops a working example from which to explicate our derivations and applications. Section II outlines the multilevel models used to estimate the average treatment effects in three-level multisite cluster-randomized trials and delineates the necessary design parameters. Section III introduces statistical power formulas. Section IV derives optimal sample calculations. Section V defines relative efficiency to be used for the comparison of designs with different sample allocations. Section VI illustrates the utility of the optimal sampling framework by comparing the sample allocations identified by the proposed framework with those of conventional frameworks. Section VII concludes the paper with a discussion.

Working Example

Before detailing the multilevel models, we provide a working substantive example to clarify presentation. We note, however, that the framework is applicable to any substantive investigation that draws on a three-level multisite cluster-randomized design. In our example, we consider the design of a three-level multisite cluster-randomized trial evaluating the impacts of a teacher development program (Cultivating Awareness and Resilience in Education or CARE; Jennings et al., 2017) on student academic achievement. The CARE program focuses on the well-being of teachers and the quality of instruction for better student outcomes. The program trains teachers on the knowledge and skills of emotional awareness, techniques and strategies for emotion regulation, and ways to apply these skills to teaching.

In our working illustration, we consider the design of a three-level multi-school teacher-randomized trial in which teachers within each school are randomly assigned to an experimental professional development group or a control group. Our analyses consider a three-level nested structure that includes: students (level one), teachers (level two), and schools (level three). The CARE program provided 30 h of in-person training plus phone coaching to teachers in the experimental group whereas teachers in the control group continued in a business-as-usual condition (e.g., Jennings et al., 2017; Institute of Education Science (IES), 2012).

The focal research question in this hypothetical study examines the degree to which exposure to the professional development program improves student reading achievement (e.g., Jennings et al., 2017; IES, 2012). In planning the evaluation, researchers must decide how to use their resources to develop a sampling plan that includes a total sample size and how that total sample will be allocated across levels in ways that promote efficiency and eventually achieve adequate power. For this reason, our working design questions probe the sample allocation plan that produces the maximum statistical power for a fixed budget (i.e., optimal design). In addition, there are often many practical constraints that preclude using an optimal sample allocation plan (e.g., limited number of students per class that can be sampled) — and for this reason we also seek to identify the sample allocation plan that maximizes power when one or more of the sample sizes at a particular level are constrained. Below we outline the corresponding models and develop a framework that addresses these core sampling questions while relaxing assumptions regarding the equality of costs across treatment conditions.

Models

Suppose a three-level multi-school teacher-randomized trial assessing the impact of CARE program on student reading achievement has a total of K schools (i.e., sites or level-three units), each school has J teachers (i.e., level-two units), and each teacher serves n students (i.e., level-one units). Let Y_{ijk} be the continuous outcome of student i in teacher j's class in school k with $i = 1, \ldots, n,\ j = 1, \ldots, J,$ and $k = 1, \ldots, K$. Let T_{jk} be the treatment indicator with $T_{jk} = 1$ for teachers in the experimental group and otherwise $T_{jk} = 0$. Further let the proportion of teachers assigned to the experimental group be p $(0 < p < 1)$ such that each school has pJ teachers in the experimental group (the CARE program) and $(1-p)J$ teachers in the control group. We can use multilevel linear models (Raudenbush & Bryk, 2002) to estimate the average treatment effects in three-level multisite cluster-randomized trails.

Although unbiased estimate of the treatment effect can be obtained without conditioning on covariates, a common design strategy is to include covariates to improve efficiency (e.g., Hedges & Hedberg, 2007a, 2007b, 2013; Jacob et al., 2010; Kelcey et al., 2016; Raudenbush et al., 2007). For this reason, let us consider $\boldsymbol{Z}_1 = (Z_{11}, \ldots, Z_{1q_1})$ as a q_1-length vector of student-level covariates with $q_1 \geq 0$ and $\boldsymbol{b}_1 = (b_{11}, \ldots, b_{1q_1})'$ be the vector of regression coefficients, the student-level or level-one model is

$$Y_{ijk} = \beta_{0jk} + \boldsymbol{b}_1 \boldsymbol{Z}_1 + \varepsilon_{ijk} \varepsilon_{ijk} \sim N\left(0, \sigma^2_{1A}\right). \tag{1}$$

Where β_{0jk} is the conditional mean score of teacher j's class in school k that varies across teachers, ε_{ijk} is the error term at the student level with a mean of zero and conditional variance of σ^2_{1A}, the subscript A represents that the variance is adjusted for covariates.

Let $\boldsymbol{Z}_2 = (Z_{21}, \ldots, Z_{2q_2})$ be a q_2-length vector of level-two (teacher-level) covariates with $q_2 \geq 0$ and $\boldsymbol{b}_2 = (b_{21}, \ldots, b_{2q_2})'$ be the vector of regression coefficients, the level-two model is

$$\beta_{0jk} = \beta_{00k} + \delta_{01k} T_{jk} + \boldsymbol{b}_2 \boldsymbol{Z}_2 + r_{0jk} \quad r_{0jk} \sim N\left(0, \sigma^2_{2A}\right). \tag{2}$$

Where β_{00k} is the conditional mean for school k, δ_{01k} is the treatment effect in school k that varies across schools, r_{0jk} is the random effect associate with teacher j in school k and it has a mean of zero and conditional variance of σ^2_{2A}.

In a same vein, let $\boldsymbol{Z}_3 = (Z_{31}, \ldots, Z_{3q})$ be a q-length vector of level-three (school-level) covariates with $q \geq 0$ and $\boldsymbol{b}_3 = (b_{31}, \ldots, b_{3q})'$, $\boldsymbol{b}_4 = (b_{41}, \ldots, b_{4q})'$ be the vectors of the regression coefficients, the site-level or school-level models are

$$\beta_{00k} = \gamma_{000} + \boldsymbol{b}_3 \boldsymbol{Z}_3 + u_{00k} u_{00k} \sim N\left(0, \sigma^2_{3A}\right), \tag{3}$$

and

$$\delta_{01k} = \delta + \boldsymbol{b}_4 \boldsymbol{Z}_3 + u_{01k} u_{01k} \sim N\left(0, \sigma^2_{\omega A}\right). \tag{4}$$

Where γ_{000} is the conditional mean across all schools, and u_{00k} is the random effect associated with school k. δ is the average treatment effect, u_{01k} is the deviance of school k from the average treatment effect and it has a mean of zero and conditional variance of $\sigma^2_{\omega A}$. $\sigma^2_{\omega A}$ is often referred as treatment-by-site variance (Raudenbush & Liu, 2000). It is assumed that the covariates included in the site-level models potentially explain some of the site-level intercept variance and treatment-by-site variance (e.g., Raudenbush & Liu, 2000).

When $q_1 = q_2 = q = 0$, the above set of equations reduces to unadjusted models without covariates. We then denote the corresponding unadjusted variance parameters at the student, teacher, school levels as σ_1^2, σ_3^2, and σ_3^2, the unadjusted treatment-by-site variance as σ_ω^2. The resulting unconditional variance partition coefficients (or intraclass correlation coefficients) at the teacher and school levels are

$$\rho_2 = \sigma_2^2/\sigma_T^2 = \sigma_2^2/(\sigma_1^2 + \sigma_2^2 + \sigma_3^2), \tag{5}$$

and

$$\rho_3 = \sigma_3^2/\sigma_T^2 = \sigma_3^2/(\sigma_1^2 + \sigma_2^2 + \sigma_3^2), \tag{6}$$

respectively. The proportions of variance at the student and teacher levels explained by covariates are

$$R_1^2 = \frac{\sigma_1^2 - \sigma_{1A}^2}{\sigma_1^2}, \tag{7}$$

and

$$R_2^2 = \frac{\sigma_2^2 - \sigma_{2A}^2}{\sigma_2^2}. \tag{8}$$

Similarly, the proportion of the treatment-by-site variance explained by covariates is

$$R_{3m}^2 = \frac{\sigma_\omega^2 - \sigma_{\omega A}^2}{\sigma_\omega^2}. \tag{9}$$

Statistical Power

When the null hypothesis is false (i.e., $\delta \neq 0$), the statistical power at the significance level α for the one-tailed test (e.g., Liu, 2003) is

$$P = 1 - H[c(\alpha, v), v, \lambda], \tag{10}$$

where $v = K - q - 1$, $c(\alpha, v)$ is the critical value in a t-distribution with v degrees of freedom and significance level of α, and $H(x, v, \lambda)$ is the cumulative distribution function of the noncentral t-distribution with v degrees of freedom and the noncentrality parameter λ. Similarly, the statistical power at the significance level α for the two-tailed test (e.g., Liu, 2003) is

$$P = 1 - H[c(\alpha/2, v), v, \lambda] + H[-c(\alpha/2, v), v, \lambda]. \tag{11}$$

The noncentrality parameter λ is defined as

$$\lambda = \frac{\delta}{\sqrt{\sigma_\delta^2}}, \tag{12}$$

with σ_δ^2 as the variance of the average treatment effect estimator. For a three-level multi-site cluster-randomized design, the variance of the average treatment effect estimator is

$$\sigma_\delta^2 = \frac{\sigma_{\omega A}^2}{K} + \frac{\sigma_{2A}^2}{K\hat{J}} + \frac{\sigma_{1A}^2}{K\hat{J}n}, \tag{13}$$

with

$$\hat{J} = [pJ(1-p)J]/[pJ + (1-p)J] = p(1-p)J. \tag{14}$$

In general, holding other factors constant, the statistical power of a design will increase when the sample size at any level gets larger. When the effect size gets larger, the statistical power will also increase.

When we standardize the treatment effect as

$$d = \frac{\delta}{\sqrt{\sigma_T^2}} = \frac{\delta}{\sqrt{\sigma_1^2 + \sigma_2^2 + \sigma_3^2}}, \tag{15}$$

the standardized treatment-by-site variance becomes

$$\omega = \frac{\sigma_\omega^2}{\sigma_1^2 + \sigma_2^2 + \sigma_3^2}. \tag{16}$$

Using the information about the variance partition coefficients and R-squared values (Equations 5 to 9), we can rewrite the variance of the standardized average treatment effect estimator as

$$\sigma_d^2 = \frac{\sigma_\delta^2}{\sigma_1^2 + \sigma_2^2 + \sigma_3^2} = \frac{p(1-p)nJ\omega\left(1 - R_{3m}^2\right) + n\rho_2\left(1 - R_2^2\right) + (1 - \rho_2 - \rho_3)\left(1 - R_1^2\right)}{p(1-p)nJK}. \tag{17}$$

The noncentrality parameter thus can be written as

$$\lambda = \frac{\delta}{\sqrt{\sigma_\delta^2}} = \frac{d}{\sqrt{\sigma_d^2}} = \frac{d\sqrt{p(1-p)nJK}}{\sqrt{p(1-p)nJ\omega\left(1 - R_{3m}^2\right) + n\rho_2\left(1-R_2^2\right) + (1 - \rho_2 - \rho_3)\left(1 - R_1^2\right)}}. \tag{18}$$

Optimal Sample Allocation

In contrast to conventional frameworks that assume equal costs between treatment conditions, we allow the costs of sampling each additional student in the experimental and control groups to differ such that c_1^T is the cost for adding a treatment student and c_1 is the cost for adding a control student. Similarly, let the costs of enrolling each additional teacher differ between the experimental and control conditions such that c_2^T is the cost for adding a treatment teacher and c_2 is the cost for adding a control teacher. Moreover, let c_3 be the costs of sampling an additional school. The total cost (m) for a three-level multi-school teacher-randomized design is then the sum of the products of marginal cost per unit and the number of units across all levels and treatment conditions. We have

$$m = K\left[pJ\left(nc_1^T + c_2^T\right) + (1 - p)J(nc_1 + c_2) + c_3\right]. \tag{19}$$

Rearranging the above cost equation, we have

$$K = \frac{m}{pJ\left(nc_1^T + c_2^T\right) + (1 - p)J(nc_1 + c_2) + c_3}. \tag{20}$$

These results describe the conventional tradeoff between sample allocation across levels — e.g., if we sample more teachers in each school, we must sample a smaller number of schools given a fixed budget. However, these results also describe a new tradeoff between sample allocation across conditions — e.g., if we sample fewer teachers in the experimental condition, we can sample more teachers in the control condition under the same budget. Yet, all combinations of sample size allocations under a fixed budget may not provide comparable statistical power. To identify an efficient sampling scheme, we need to consider the information about sampling costs within the context of the power formula (e.g., λ or σ_d^2). Substituting the above equation for K in Equation 17, we have the variance of standardized average treatment effect estimator as

$$\sigma_d^2 = \frac{\left[p(1-p)nJ\omega\left(1-R_{3m}^2\right) + n\rho_2\left(1-R_2^2\right) + (1-\rho_2-\rho_3)\left(1-R_1^2\right)\right]\left[pJ\left(nc_1^T + c_2^T\right) + (1-p)J(nc_1 + c_2) + c_3\right]}{p(1-p)nJm}. \quad (21)$$

Prior literature has assumed equal costs across treatment conditions (i.e., $c_1^T = c_1$ and $c_2^T = c_2$) and derived the optimal sampling values by minimizing the variance of the treatment effect estimator, which is approximately equivalent to maximizing the statistical power under a fixed cost (e.g., Raudenbush, 1997). We used a similar approach to identify the optimal sample allocation while relaxing the assumption of equal costs across treatment conditions. The results suggested that readily interpretable expressions can be obtained for the optimal student-level sample size (n) and optimal teacher-level sample size (J), but not for the optimal proportion of teachers assigned to the experimental condition (p). The derivation process for these equations is presented in Appendix.

To get a sense of what drives the optimal sampling plan, we outline the solutions for the optimal number of students per teacher (n) and optimal number of teachers per school (J) for fixed values of the remaining parameters. In terms of the optimal number of students per teacher (n), our results show that

$$n = \sqrt{\frac{(1-\rho_2-\rho_3)(1-R_1^2)}{p(1-p)J\omega(1-R_{3m}^2) + \rho_2(1-R_2^2)} \times \frac{(1-p)Jc_2 + pJc_2^T + c_3}{(1-p)Jc_1 + pJc_1^T}}. \quad (22)$$

The resulting optimal n parallels the results found in conventional frameworks that are conceptually composed of two primary components: the ratio of the error variances across levels and the cost ratio across levels. Specifically, the first component in our result approximately captures the ratio between the level-one (student-level) conditional variance and the conditional variance of the average treatment effect (i.e., conditional variance of the effect across schools and conditional variance across teachers within schools). The second component roughly captures the ratio between the (weighted) summative cost of sampling upper-level units (teachers and schools) and the (weighted) cost of sampling level-one units (students). When assembled, the formula conceptually expands on the results found in conventional frameworks by delineating the roles and relative contributions of competing costs. By incorporating differential costs, the results introduce a flexible weight parameter (p) while conventional frameworks can be viewed as implicitly fixing the value of that weight parameter to .5 because of the equal costs assumption.

Our analyses found a similar result for the optimal number of teachers per school (J). The optimal J is

$$J = \sqrt{\frac{n\rho_2(1-R_2^2) + (1-\rho_2-\rho_3)(1-R_1^2)}{n\omega(1-R_{3m}^2)} \times \frac{c_3}{(1-p)(c_1 n + c_2) + p\left(c_1^T n + c_2^T\right)} \times \frac{1}{p(1-p)}}. \quad (23)$$

Again, this result suggests that the optimal number of teachers per school (J) is guided by the product of the cross-level ratios of variance and cost, and additionally by a third component in which the denominator is the product of sampling weights across treatment conditions ($\frac{1}{p(1-p)}$). This third component has a minimum value of 4 when $p = .5$ or in balanced designs. The result implies that once optimal sampling proportion (p, which will be discussed later) deviates from .5, we will have larger J values than those in a conventional framework with $p = .5$.

The resulting formulas indicate that the optimal p, n and J are functions of two sets of well-known parameters: variance partition parameters and cost structure parameters. The variance partition parameters include variance partition coefficients at different levels (ρ_2, ρ_3, and ω) and proportions of variance explained by covariates (R_1^2, R_2^2, and R_{3m}^2). Cost structure parameters include the costs of sampling a unit across levels and conditions (c_1, c_1^T, c_2, c_2^T, and c_3).

The equation that identifies the optimal proportion of clusters/teachers assigned to treatment (p) is outlined in the Equation A1 in Appendix. In the Illustration section, we survey a range of parameter values to explore what factors impact the optimal p. In the general case, we can numerically solve for the roots of n, J, p simultaneously using Equations 22, 23, and A1 in the Appendix (Shen & Kelcey, 2020c). These solutions have been implemented in the *R* package *odr* (Shen & Kelcey, 2020a). We denote the level-specific and condition-specific optimal sample using n^o, J^o, and p^o. Once p^o, n^o, and J^o are identified, the corresponding optimal number of sites/schools (K^o) can be determined by (a) the total budget through Equation 20 when the budget is fixed or (b) by a targeted statistical power level through Equation 11 (along with Equations 12, 13, and 14) when power is preset. All these types of power and sample size calculations have been implemented in the *R* package *odr* (Shen & Kelcey, 2020a).

Implications

As outlined above, the optimal sampling values across levels are mainly driven by two components that are consistent with prior literature: the conditional variance ratios across levels and the sampling cost ratios across levels. The optimal sample size at the level of randomization (e.g., teacher level) is additionally impacted by the product of sampling weights between treatment conditions. The implication of the first component (variance ratios) is that optimal sampling is positively proportional to the conditional variance decomposition — that is, holding other factors constant, increases in the (relative) conditional variance at a level warrant increases in the sample size for that level.

For example, when the ratio of conditional variance at the student level to conditional variance at the teacher and school levels $\left(\frac{(1-\rho_2-\rho_3)(1-R_1^2)}{p(1-p)J\omega(1-R_{3m}^2)+\rho_2(1-R_2^2)}\right)$ gets larger, the optimal n tends to be larger (see Equation 22). Holding other factors constant, larger (residual) variances at the teacher and school levels relative to the conditional student-level variance typically suggest that it is beneficial to sample a larger number of teachers and schools in exchange for fewer students per teacher. The optimal teacher-level sample size follows a similar pattern but splits the pertinent variance ratio into a numerator that contains the conditional teacher- and student-level variation and a denominator that contains school-level variation (Equation 23). Once again, the implication is that holding other factors constant, an increase in the conditional teacher-level variance suggests sampling more teachers per school in exchange for fewer students per teacher.

The implication of second component (cost ratios) is similar but opposite — optimal sampling for a level is inversely proportional to the cost of sampling a unit at that level. That is, intuitively, the optimal sampling strategy suggests sampling additional (relatively) inexpensive units in exchange for fewer expensive units under a fixed budget.

For the optimal number of students, as the ratio of the marginal costs at the school and teacher levels to the student-level costs $\left(\frac{(1-p)Jc_2 + pJc_2^T + c_3}{(1-p)Jc_1 + pJc_1^T}\right)$ grows, the optimal n increases. When the ratio of the marginal costs at the school level to those of the teacher and student levels $\left(\frac{c_3}{(1-p)(c_1 n + c_2) + p\left(c_1^T n + c_2^T\right)}\right)$ grows, the optimal J increases. Again, under a fixed budget, the optimal sampling strategy is to sample additional inexpensive units while dropping some of the expensive units.

As we discussed before, the optimal sample size at the teacher level (J) is additionally impacted by the value of optimal p. In turn, optimal p values are mainly impacted by sampling cost ratio between treatment conditions (see the Illustration section). That is, the optimal number of teachers per school is also indirectly impacted by the sampling cost ratio between treatment conditions that is transferred through the optimal p value. One implication of these relationships is that when we fix p (as in conventional frameworks), analyses will typically reach a different and sub-optimal value.

Collectively, the variance and cost ratios (including cost ratios across levels and treatment conditions) reflect two considerations that are well-known across multilevel experimental design. The variance ratio component suggests that a relatively larger conditional variance at a given level necessitates a relatively bigger sample size at that level. This result reaffirms and expands upon the well-known strategy that levels with larger residual variance will typically need larger sample sizes to attain a targeted level of precision. Equally, the cost ratio component suggests that relatively cheaper marginal costs of sampling a unit at a level translate into larger sample sizes at that level. That is, when the cost of sampling a unit at a particular level is relatively lower than other levels, it is prudent to oversample that level because doing so increases precision while incurring little relative cost. The same principle applies to the cost ratio between treatment conditions — optimal design analyses suggest sampling a larger portion of units assigned in the less expensive condition.

In contrast to the optimal sample sizes at the student and teacher levels, the roles of parameters in determining the optimal proportion of teachers to be assigned to the treatment condition (p^o) are less clear in Equation A1. In particular, the results of our analysis indicated that the optimal proportion of teachers assigned to treatment is governed by a complex interplay of the concomitant parameters (see Appendix). Moreover, this result is not unique to multisite cluster-randomized designs but rather extends to many types of multilevel designs (Shen & Kelcey, 2020b, 2020c). However, recent studies of optimal sampling in other types of multilevel experiments suggested that the optimal p is disproportionally driven by the sampling cost ratio between treatment conditions (Shen & Kelcey, 2020b, 2020c). For this reason, the optimal p for three-level multisite cluster-randomized studies may also be primarily dominated by the cost ratio between treatment conditions. We subsequently surveyed this hypothesis using different sets of parameter values (see the Illustration section).

Constrained Optimal Sample Allocation

Equations 22, 23 and A1 jointly provide optimal sampling solutions when each of the three sample sizes are mutable. However, there may be practical reasons that constrain one or more of the sample sizes to be fixed or limited to certain values. For example, schools may only have a maximum of four teachers in each grade. In this situation, the number of teachers per school will be limited to a maximum of four. Under such practical constraints, a constrained optimal sampling plan can be obtained by substituting the constrained value into an appropriate equation (i.e., one of Equations 22, 23 and A1) and solving for the remaining (unconstrained) parameters. Solutions for all possible combinations of constrained and unconstrained optimal sampling strategies have been implemented in the *R* package *odr* (Shen & Kelcey, 2020a).

The constrained framework also enables us to establish a theoretical connection between the proposed framework and the conventional framework for three-level multisite cluster-randomized trails. Specifically, when placing constraints on the per unit sampling costs across treatment conditions and requiring balanced assignment (i.e., $p = .5$), our proposed framework reduces to the optimal sampling plans outlined in conventional frameworks (e.g., Hedges & Borenstein, 2014; Raudenbush, 1997; Raudenbush & Liu, 2000). To illustrate this, let us assume a balanced design ($p = .5$) and set the cost of sampling a unit at each level be coarsely summarized by the average cost across conditions such that: (a)$C_1 = pc_1^T + (1-p)c_1$ and (b) $C_2 = pc_2^T + (1-p)c_2$. That is, C_1 and C_2 represent the respective average costs of sampling a student and a teacher in both treatment conditions (i.e., $C_1 = \frac{c_1^T + c_1}{2}$ and $C_2 = \frac{c_2^T + c_2}{2}$). In this scenario, those constraints produce the following constrained optimal design parameter expressions:

$$n = 2\sqrt{\frac{(1-\rho_2-\rho_3)(1-R_1^2)}{J\omega(1-R_{3m}^2) + 4\rho_2(1-R_2^2)} \times \left[\frac{C_2}{C_1} + \frac{c_3}{C_1 J}\right]}, \tag{24}$$

$$J = 2\sqrt{\frac{n\rho_2(1-R_2^2) + (1-\rho_2-\rho_3)(1-R_1^2)}{n\omega(1-R_{3m}^2)} \times \frac{c_3}{C_1 n + C_2}}. \tag{25}$$

The above constrained formulas produce the same values for n and J as those of the conventional optimal sampling framework for three-level multisite cluster-randomized trials (Hedges & Borenstein, 2014). Equations 24 and 25 are in a different format from those in the conventional approach because of differences in notations for sample sizes and the method of standardizing the treatment-by-site variance. For example, Hedges & Borenstein (2014) denoted the level-two sample size as the number of clusters per site in each treatment condition rather than the total number of clusters per site for both treatment conditions as we did. Also, the treatment-by-site variance was standardized by scaling the unconditional site-level within treatment variance as one or $\omega = \frac{\sigma_\omega^2}{\sigma_3^2}$ (Hedges & Borenstein, 2014).

Relative Efficiency

When planning multilevel experiments, designs with different sample allocations are typically considered. Our proposed framework helps structure and start those deliberations by identifying the most powerful sample allocation for a given budget. However, as noted above, in many instances researchers may discuss sub- or semi-optimal designs that balance efficient sampling with practical design considerations. In such cases, it can be helpful to compare the relative efficiency of two or more potential designs to understand the efficiency lost by sub-optimal sampling.

We use σ_d^2 to denote the sampling variance for a suboptimal or alternative sampling plan and $\sigma_d^2 o$ to denote the sampling variance for a more optimal (efficient) sampling plan. The relative efficiency is $\sigma_d^2 o/\sigma_d^2$ (e.g., Korendijk et al., 2010). Using the information in Equation 21, the relative efficiency between two sampling plans for a three-level multisite experiment is

$$RE = \frac{\sigma_d^2 o}{\sigma_d^2} = \frac{p^o(1-p^o)n^oJ^o\omega(1-R_{3m}^2) + n^o\rho_2(1-R_2^2) + (1-\rho_2-\rho_3)(1-R_1^2)}{p(1-p)nJ\omega(1-R_{3m}^2) + n\rho_2(1-R_2^2) + (1-\rho_2-\rho_3)(1-R_1^2)} \times \frac{\left[(1-p^o)(c_1n^oJ^o + c_2J^o) + p^o\left(c_1^Tn^oJ^o + c_2^TJ^o\right) + c_3\right]p(1-p)nJ}{\left[(1-p)(c_1nJ + c_2J) + p\left(c_1^TnJ + c_2^TJ\right) + c_3\right]p^o(1-p^o)n^oJ^o}. \quad (26)$$

Where p^o, n^o, and J^o represent the optimal design parameters in a more efficient sampling plan (e.g., the solved values of optimal design parameters expressed in Equations 22, 23 and A1). p, n, and J represent the values used in a sub-optimal or alternative sampling plan. We have implemented this formula in the *R* package *odr* (Shen & Kelcey, 2020a) to assist in the judicious comparison of different designs.

Literature suggests that an RE of .90 or above is considered good and RE values between .80 and .90 are considered acceptable (Korendijk et al., 2010). Beyond the relative variance ratio between two designs, RE also has implications on sample size and budget. Holding other parameters constant, the variance of the standardized average treatment effect estimator (σ_d^2) and the number of sites/schools (K) are inversely proportional (see Equation 17). From a sample size perspective, a sub-optimal design would need (1 − RE)/RE × 100% more schools than an optimal design to reach a comparable error variance at the sub-optimal sample allocation (i.e., values of p, n, and J). For example, a design with an RE value of .80 would need to sample 25% (i.e., (1 − .80)/.80) additional schools than an optimal design to achieve a same design precision that the optimal design produces.

Notice that the additional percentage in the number of sites (i.e., (1-RE)/RE × 100%) is also the additional percentage in budget requested by a sub-optimal design to achieve a same error variance level. This is because, holding other parameters constant, the budget (m) is directly proportional to the number of sites (see Equation 19). That is, a design with an RE value of .80 would need 25% additional budget to maintain the same design precision with an optimal design.

Alternatively, prior literature has often contrasted designs using their minimum detectable effect sizes (e.g., Bloom, 1995). We can contrast the minimum detectable effect size under two sampling designs by noting that

$$d^o = d\sqrt{RE}. \quad (27)$$

Here we use d^o and d to denote the respective standardized minimum detectable effect sizes for the optimal and sub-optimal design with RE as their relative efficiency.

Illustration

Next, we survey the resulting optimal sampling plans within the context of our working example by considering a range of cost structures and design parameters taken in literature (e.g., Jennings et al., 2017; IES, 2012). Our illustrations (a) outline how cost structures and values of concomitant parameters (e.g., intraclass correlation coefficients, treatment-by-site variance) impact optimal sampling plans and (b) compare the sample size plans identified under our framework and those of the conventional framework that assume equal cross-treatment costs.

For the intraclass correlation coefficients and R-squared values, we use the empirical values reported in a recent study for reading achievement as a starting point and then vary these values. The intraclass correlation coefficients at the school and teacher levels reported in a recent literature are .06 and .09, respectively (i.e., $\rho_3 = .06$ and $\rho_2 = .09$; Jacob et al., 2010). Similarly, we set proportions of variance at the student and teacher levels explained by covariates (e.g., pretest scores) as .50, which are more conservative estimations than the empirical values reported in literature (Jacob et al., 2010).

To see how variations in the intraclass correlation coefficients potentially impact optimal sampling plans, we consider three other combinations of intraclass correlation coefficients that represent a high proportion of variance at the teacher level ($\rho_3 = .04$, $\rho_2 = .20$), a high proportion of variance at the school level ($\rho_3 = .20,\ \rho_2 = .04$), and high proportions of variance at both the teacher and school levels ($\rho_3 = .20,\ \rho_2 = .20$).

Similarly, drawing on our working example, we consider a setting in which costs are roughly equal across conditions for sampling one additional student ($c_1 = c_1^T$), but costs are unequal across conditions when sampling one additional teacher ($c_2 \neq c_2^T$). Specifically, we let the cost of sampling one additional student be US\$10 regardless of which treatment condition a student is in (i.e., $c_1 = c_1^T =$ US\$10), the cost of sampling one teacher in the control condition be US\$50 (i.e., $c_2 =$ US\$50), the cost of sampling one teacher in the experimental group be US\$3,000 or US\$6,000 (i.e., $c_2^T =$ US\$3,000 or $c_2^T =$ US\$6,000; to reflect the additional costs for training and mentoring), and the cost of sampling one additional school be US\$1,000 or US\$2,000 (i.e., $c_3 =$ US\$1,000 or $c_3 =$ US\$2,000; to reflect potential costs for recruiting a school and/or incentive pay to a school).

Furthermore, we considered a design targeting an average treatment effect as small as 0.2 ($d = 0.2$) with a treatment-by-site variance of 0.01 ($\omega = 0.01$) or 0.04 ($\omega = 0.04$). Finally, we considered 30% of treatment-by-site variance to be explained by covariates (i.e., $R_{3m}^2 = .30$). In sum, our illustration considered 32 conditions (i.e., a factorial design of 4 cost structures, 4 intraclass correlation coefficients, and 2 treatment-by-site variances). We report the resulting optimal sample sizes in the middle four columns of Table 1 (i.e., the columns of $p^o,\ n^o,\ J^o,\ K^o$).

The results suggest that two primary factors determine the value of optimal p^o. The first factor is the cost ratio between treatment conditions at the individual- and cluster-level. Specifically, for same intraclass correlation coefficients and treatment-by-site variance, the larger the cost ratio between treatment and control conditions (e.g., c_2^T/c_2 or $\frac{c_1^T n + c_2^T}{c_1 n + c_2}$) is, the smaller the optimal p value (i.e., p^o). Holding other parameters equal, when the cost of sampling a school rises, p^o essentially does not change. That is, the optimal p is mainly determined by the cost ratio at the level of randomization (e.g., teacher-level in this example). The second factor impacting the value of optimal p is the intraclass correlation coefficients at the teacher- and school-levels. For example, increases in the variance at the teacher level results in a smaller p^o (see the first and fourth combinations of intraclass correlation coefficients versus the other two combinations). The change of treatment-by-site variance has a negligible effect on p^o as its values are identical for $\omega = 0.01$ and $\omega = 0.04$.

Table 1. Optimal design parameters in three-level multi-school teacher-randomized trials

		Proposed framework				Conventional framework				
(ρ_3, ρ_2)	(c_2^T, c_3) (US$)	p^o	n^o	J^o	K^o	$n_{p=.5}$	$J_{p=.5}$	$K_{p=.5}$	Power	RE
$\omega = 0.01$										
(.04, .20)	(3,000, 1,000)	.20	15.74	11.62	16.19	24.07	6.12	18.66	.69	.78
	(3,000, 2,000)	.20	15.74	16.43	13.09	24.07	8.66	13.69	.68	.78
	(6,000, 1,000)	.17	19.91	9.92	20.04	33.90	4.35	25.26	.66	.73
	(6,000, 2,000)	.17	19.91	14.03	15.37	33.90	6.15	18.06	.66	.73
(.06, .09)	(3,000, 1,000)	.24	26.61	6.88	12.94	37.95	4.11	14.59	.71	.82
	(3,000, 2,000)	.24	26.61	9.73	10.24	37.95	5.81	11.02	.71	.83
	(6,000, 1,000)	.20	34.14	5.72	15.31	53.45	2.92	17.39	.68	.77
	(6,000, 2,000)	.20	34.14	8.08	12.44	53.45	4.12	13.99	.69	.78
(.20, .04)	(3,000, 1,000)	.27	39.86	4.18	11.01	53.83	2.74	10.75	.71	.86
	(3,000, 2,000)	.27	39.86	5.91	8.68	53.83	3.87	9.07	.73	.87
	(6,000, 1,000)	.23	51.67	3.40	13.67	75.81	1.94	13.41	.69	.81
	(6,000, 2,000)	.23	51.67	4.81	9.78	75.81	2.75	10.24	.70	.82
(.20, .20)	(3,000, 1,000)	.20	13.74	11.99	15.95	21.39	6.12	18.46	.68	.77
	(3,000, 2,000)	.20	13.74	16.96	12.38	21.39	8.66	13.56	.68	.77
	(6,000, 1,000)	.16	17.33	10.31	20.66	30.12	4.35	25.03	.65	.72
	(6,000, 2,000)	.16	17.33	14.58	15.01	30.12	6.15	17.90	.66	.73
$\omega = 0.04$										
(.04, .20)	(3,000, 1,000)	0.20	15.74	5.81	32.82	24.07	3.06	37.82	.70	.79
	(3,000, 2,000)	0.20	15.74	8.22	26.53	24.07	4.33	30.27	.71	.80
	(6,000, 1,000)	0.17	19.91	4.96	40.61	33.90	2.17	51.12	.67	.74
	(6,000, 2,000)	0.17	19.91	7.01	31.18	33.90	3.07	36.60	.68	.75
(.06, .09)	(3,000, 1,000)	0.24	26.61	3.44	29.33	37.95	2.05	29.59	.72	.84
	(3,000, 2,000)	0.24	26.61	4.87	20.67	37.95	2.90	22.29	.73	.85
	(6,000, 1,000)	0.20	34.14	2.86	31.04	53.45	1.46	49.10	.70	.79
	(6,000, 2,000)	0.20	34.14	4.04	25.20	53.45	2.06	28.36	.71	.80
(.20, .04)	(3,000, 1,000)	0.27	39.86	2.09	22.27	53.83	1.37	28.77	.74	.88
	(3,000, 2,000)	0.27	39.86	2.96	17.43	53.83	1.94	18.24	.75	.89
	(6,000, 1,000)	0.23	51.67	1.70	22.71	75.81	0.97	27.18	.73	.83
	(6,000, 2,000)	0.23	51.67	2.40	22.71	75.81	1.37	27.18	.73	.85
(.20, .20)	(3,000, 1,000)	0.20	13.74	6.00	32.35	21.39	3.06	37.42	.70	.78
	(3,000, 2,000)	0.20	13.74	8.48	26.18	21.39	4.33	29.97	.70	.80
	(6,000, 1,000)	0.16	17.33	5.16	41.86	30.12	2.17	50.66	.67	.73
	(6,000, 2,000)	0.16	17.33	7.29	32.07	30.12	3.07	36.30	.67	.75

Note. The above calculations are based on $c_1 = c_1^T =$ US\$10, $c_2 =$ US\$50, $q = 1$, $R_1^2 = R_2^2 = .50$, and $R_{3m}^2 = .30$. ρ_2 and ρ_3 are the intraclass correlation coefficients at the teacher and school levels, respectively. p^o, n^o, and J^o are the optimal design parameters identified by the proposed framework, K^o is the number of schools/sites required to achieve 80% power under the optimal sample allocation. $n_{p=.5}$ and $J_{p=.5}$ are the constrained optimal sample sizes identified by the conventional framework (Hedges & Borenstein, 2014) with a constrained $p = .5$, $K_{p=.5}$ is the number of schools/sites the conventional framework can sample under the same budget. Power is the statistical power that a conventional optimal sampling plan can achieve under the same budget required by the proposed sampling plan to produce 80% power. RE is the relative efficiency of the design identified in the conventional framework compared with the corresponding one in the proposed framework.

These findings echo those found in cluster-randomized trials and two-level multisite randomized trials (Shen & Kelcey, 2020b, 2020c).

As for the values of n^o and J^o, the trends are aligned with their respective equations and those implications we have outlined in a previous section: n^o and J^o are impacted by both sampling cost and conditional variance ratios across levels. Holding other factors fixed, when sampling a teacher is more expensive, the optimal number of students (n^o) gets larger; when the (conditional) variance at the teacher and school levels gets larger, the value of n^o is smaller. Adjustments to the treatment-by-site variance do not appear to impact the value of n^o (e.g., the results are identical for the two ω values; see Table 1).

In contrast, the value of J^o is impacted by changes in the treatment-by-site variance in our examples. A larger treatment-by-site variance necessitates more sites/schools to clearly separate out the variance attributable to effect heterogeneity and sampling error. As a result, when treatment-by-site variance grows, smaller values prevail for J^o. The value of J^o is also impacted by the cost ratio across levels — when the cost of sampling a school/site increases, sampling a teacher is relatively cheaper, resulting in a larger value of J^o (Table 1).

Comparison with the Conventional Framework

Using the same 32 conditions, we compare the optimal sampling plans under the proposed framework with those of the conventional framework (Hedges & Borenstein, 2014). Our analysis for the conventional framework assumes that the proportion of teachers assigned to the treatment is $p = .5$, and that costs of sampling a control unit and a treated unit at each level are equal to the average costs of treatment and control units at that level. For example, when the true costs of sampling a teacher in a control and treatment condition are $c_2 = \text{US\$50}$ and $c_2^T = \text{US\$3,000}$, we use their average costs ($C_2 = \frac{c_2^T + c_2}{2} = \frac{\text{US\$3,000} + \$50}{2} = \text{US\$1,525}$) for the conventional framework. For the conventional framework, we present the optimal sample size values along with the relative efficiency comparing to designs identified by the proposed framework.

For comparison purposes, we also report the total number of schools to achieve 80% power at the optimal sample allocation identified by the proposed framework (the K^o column in Table 1). Thus, a total cost can be calculated (e.g., through Equation 19) to achieve 80% power for an optimal design. Because the conventional optimal sampling framework identifies different sample sizes at the student and teacher levels (the $n_{p=.5}$ and $J_{p=.5}$ columns in Table 1), for the same total cost we can subsequently calculate the total number of schools can be sampled at the conventional optimal size allocation (the $K_{p=.5}$ column in Table 1). Given all these parameter values, we report the statistical power a conventional optimal design can achieve (i.e., the Power column in Table 1) under the same resources where 80% power has been achieved by the optimal sampling plans in the proposed framework.

The results for the conventional optimal sampling framework are presented in the right four columns in Table 1 (i.e., columns of $n_{p=.5}$, $J_{p=.5}$, RE, and Power). The primary result is that our framework consistently suggests sampling more teachers in favor of fewer students and schools when it is more costly to sample treatment teachers than control teachers, which is aligned with our prior explanation of optimal sample size formulas. The intuition motivating this strategy is that because control teachers demand relatively fewer resources, it is often more efficient to oversample control teachers.

More generally, there tends to be a large initial reduction in the error variance of the treatment effect associated with sampling additional control teachers relative to other units (schools, students, treatment teachers). However, the accelerated reduction of the error variance associated with oversampling control teachers eventually levels off such that it becomes more efficient to sample other units (i.e., treatment teachers, students, and schools). In contrast, the conventional framework overlooks this critical opportunity because it considers conditions and units exchangeable in terms of cost.

Our results also suggest that the differences in sampling plans selected by the frameworks have nontrivial impacts on relative efficiency and power (Table 1). Across the conditions sampled, the average relative efficiency of designs in the conventional framework is only 80%. That is, designs identified by the conventional framework need an average of 25% additional budget to sample 25% more schools so that these designs can produce the same power level of 80%.

Similarly, across the conditions sampled, the results suggest that the average power produced by the proposed optimal sampling framework is 15% larger than that of the conventional optimal sampling framework. Under identical resources, the proposed framework can generate 15% more power than the conventional framework just by nominally modifying the sample allocation. Collectively, these findings also echo the nonlinear relationship between power and sample size. Thus, the RE value between two designs will be different from their relative power (e.g., Cox & Kelcey, 2019). To gain an average of 15% more power, we need to sample an average of 25% more schools, which is substantially larger than the 15% increase on power.

Discussion

In this article, we develop a more flexible optimal sampling framework for three-level multisite group-randomized trials. This development parallels those addressing sampling and planning issues in other types of multiple designs (e.g., Shen & Kelcey, 2020b, 2020c). The proposed framework expands the scope of prior research on optimal sampling that assumed equal sampling costs between treatment conditions (e.g., Hedges & Borenstein, 2014) by operationalizing more flexible cost structures of sampling across conditions and levels.

We find that when costs vary across treatment conditions, the proposed framework can often identify more efficient and more powerful sampling plans relative to the conventional framework that assumed equal costs between treatment conditions with fixed sample allocation between conditions. In part, our results substantiate and replicate well-known design strategies. For example, our results suggest that an efficient strategy is to sample units (e.g., students, teachers, or schools) proportional to their relative cost and variance. In other ways, however, our results expand upon these strategies. For instance, when units differ in their costs across treatment conditions, our results suggest that oversampling units in the lower cost condition (e.g., control condition) can provide a strong return on investment (e.g., more efficient and powerful designs).

In this way, the results of our study potentially improve upon theoretical benchmarks for optimal sampling guidelines. The resulting formulas and guidelines we developed capture a critical starting point for design conversations and decisions regarding sampling considerations. In most empirical studies, practical considerations outside the scope of conventional experimental design parameters will additionally influence the eventual sampling plan. The results from our framework offer more efficient and nuanced starting points for researchers to balance theoretical and practical considerations.

Like the conventional framework, the added value of our framework is also contingent upon the plausibility of concomitant parameter values used as inputs (e.g., intraclass correlation coefficients). The literature has compiled empirical values for a variety of design parameters (Hedges & Hedberg, 2007a, 2007b, 2013; Kelcey et al., 2016; Raudenbush et al., 2007), including those can be used for planning three-level multi-school teacher-/class-randomized trials that are similar to our example (Dong et al., 2016; Jacob et al., 2010; Zhu et al., 2012). However, compared with conventional optimal sampling frameworks and power analysis, our framework requires additional information about the approximate costs of sampling treatment and control units. In some instances, such information can be roughly estimated from publicly available information at cost centers (e.g., CostOut at https://www.cbcsecosttoolkit.org/). Alternatively, in the absence of empirical evidence as to the parameter values, a common approach is to conduct pilot studies and/or consider a range of plausible values for parameters to understand the implications, limitations, and sensitivity of different sampling plans.

The proposed framework can be applied to a diverse range of studies beyond the substantive examples provided (i.e., three-level multi-school teacher-randomized trials). For example, a common design in settings of students nested within schools nested within districts draws on a three-level multi-district school-randomized study to target student achievement gains from whole school reform programs (e.g., Fahle & Reardon, 2018; Hedberg & Hedges, 2014; Westine, 2016). Investigators can find empirical values of intraclass correlation coefficients in the three-level nesting for different achievement domains (e.g., Fahle & Reardon, 2018; Hedberg & Hedges, 2014; Hedges & Hedberg, 2013; Kelcey & Shen, 2017; Westine, 2016) to help design efficient three-level experimental studies using the framework presented in this article.

The optimal sampling framework developed in this article offers investigators the theory and tools to incorporate disparate costs when designing studies to detect main effects. Although research into main effects represents a principal estimand in many studies, recent literature has emphasized the value in expanding such questions with those that probe effect heterogeneity and theories of action through investigations of moderation and mediation. The framework developed in this study could be extended to questions of mediation and/or moderation effects.

Open Scholarship

This article has earned the Center for Open Science badge for Open Materials through Open Practices Disclosure. The materials are openly accessible at https://osf.io/qamkx/. To obtain the author's disclosure form, please contact the Editor.

Acknowledgments

We thank the editor, Dr. Luke Miratrix, and three anonymous reviewers for their helpful comments and suggestions on the earlier draft that led to substantial improvement of the manuscript.

Disclosure statement

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

This paper is based in part on work supported by the National of Academy of Education (NAEd) and Spencer Foundation through the NAEd/Spencer Dissertation Fellowship awarded to the first author.

ORCID

Zuchao Shen http://orcid.org/0000-0003-3483-0451

References

Bloom, H. S. (1995). Minimum detectable effects: A simple way to report the statistical power of experimental designs. *Evaluation Review*, *19*(5), 547–556. https://doi.org/10.1177/0193841X9501900504

Cox, K., & Kelcey, B. (2019). Robustness of statistical power in group-randomized studies of mediation under an optimal sampling framework. Methodology, *15*(3), 106–118. https://doi.org/10.1027/1614-2241/a000169

Dong, N., Reinke, W. M., Herman, K. C., Bradshaw, C. P., & Murray, D. W. (2016). Meaningful effect sizes, intraclass correlations, and proportions of variance explained by covariates for planning two- and three-level cluster randomized trials of social and behavioral outcomes. *Evaluation Review*, *40*(4), 334–377. https://doi.org/10.1177/0193841X16671283

Fahle, E. M., & Reardon, S. F. (2018). How much do test scores vary among school districts? New estimates using population data, 2009–2015. *Educational Researcher*, *47*(4), 221–234. https://doi.org/10.3102/0013189X18759524

Greenleaf, C. L., Litman, C., Hanson, T. L., Rosen, R., Boscardin, C. K., Herman, J., Schneider, S. A., Madden, S., Jones, B., & Jones, B. (2011). Integrating literacy and science in biology: Teaching and learning impacts of reading apprenticeship professional development. *American Educational Research Journal*, *48*(3), 647–717. https://doi.org/10.3102/0002831210384839

Hedberg, E. C., & Hedges, L. V. (2014). Reference values of within-district intraclass correlations of academic achievement by district characteristics: Results from a meta-analysis of district-specific values. *Evaluation Review*, *38*(6), 546–582. https://doi.org/10.1177/0193841X14554212

Hedges, L. V., & Hedberg, E. C. (2007a). Intraclass correlation values for planning group-randomized trials in education. *Educational Evaluation and Policy Analysis*, *29*(1), 60–87. https://doi.org/10.3102/0162373707299706

Hedges, L. V., & Hedberg, E. C. (2007b). Intraclass correlations for planning group randomized experiments in rural education. *Journal of Research in Rural Education*, *22*(10), 1–15.

Hedges, L. V., & Hedberg, E. C. (2013). Intraclass correlations and covariate outcome correlations for planning two- and three-level cluster-randomized experiments in education. *Evaluation Review*, *37*(6), 445–489. https://doi.org/10.1177/0193841X14529126

Hedges, L. V., & Borenstein, M. (2014). Conditional optimal design in three-and four-level experiments. *Journal of Educational and Behavioral Statistics*, *39*(4), 257–281. https://doi.org/10.3102/1076998614534897

Hiscock, H., Bayer, J. K., Price, A., Ukoumunne, O. C., Rogers, S., & Wake, M. (2008). Universal parenting programme to prevent early childhood behavioural problems: Cluster randomised trial. *BMJ (Clinical Research Ed.)*, *336*(7639), 318–321. https://doi.org/10.1136/bmj.39451.609676.AE

Institute of Education Science (IES). (2012). *Improving classroom learning environments by cultivating awareness and resilience in education (CARE): A cluster randomized controlled efficacy trial*. https://ies.ed.gov/funding/grantsearch/details.asp?ID=1233.

Jacob, R., Goddard, R., Kim, M., Miller, R., & Goddard, Y. (2015). Exploring the causal impact of the McREL balanced leadership program on leadership, principal efficacy, instructional climate, educator turnover, and student achievement. *Educational Evaluation and Policy Analysis*, *37*(3), 314–332. https://doi.org/10.3102/0162373714549620

Jacob, R., Zhu, P., & Bloom, H. (2010). New empirical evidence for the design of group randomized trials in education. *Journal of Research on Educational Effectiveness*, *3*(2), 157–198. https://doi.org/10.1080/19345741003592428

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

Jennings, P. A., Brown, J. L., Frank, J. L., Doyle, S., Oh, Y., Davis, R., Rasheed, D., DeWeese, A., DeMauro, A. A., Cham, H., & Greenberg, M. T. (2017). Impacts of the CARE for teachers program on teachers' social and emotional competence and classroom interactions. *Journal of Educational Psychology*, *109*(7), 1010–1028. https://doi.org/10.1037/edu0000187

Kelcey, B., & Shen, Z. (2017). Planning community-based assessments of HIV educational intervention programs in sub-Saharan Africa. *Health Education Research*, *32*(4), 353–363. https://doi.org/10.1093/her/cyx046

Kelcey, B., Shen, Z., & Spybrook, J. (2016). Intraclass correlation coefficients for designing cluster-randomized trials in sub-Saharan Africa education. *Evaluation Review*, *40*(6), 500–525. https://doi.org/10.1177/0193841X16660246

Konstantopoulos, S. (2009). Incorporating cost in power analysis for three-level cluster-randomized designs. *Evaluation Review*, *33*(4), 335–357. https://doi.org/10.1177/0193841X09337991

Konstantopoulos, S. (2011). Optimal sampling of units in three-level cluster randomized designs: An ANCOVA framework. *Educational and Psychological Measurement*, *71*(5), 798–813. https://doi.org/10.1177/0013164410397186

Korendijk, E. J., Moerbeek, M., & Maas, C. J. (2010). The robustness of designs for trials with nested data against incorrect initial intracluster correlation coefficient estimates. *Journal of Educational and Behavioral Statistics*, *35*(5), 566–585. https://doi.org/10.3102/1076998609360774

Liu, X. (2003). Statistical power and optimum sample allocation ratio for treatment and control having unequal costs per unit of randomization. *Journal of Educational and Behavioral Statistics*, *28*(3), 231–248. https://doi.org/10.3102/10769986028003231

Mosteller, F. (1995). The Tennessee study of class size in the early school grades. *The Future of Children*, *5*(2), 113–127. https://doi.org/10.2307/1602360

Raudenbush, S. W. (1997). Statistical analysis and optimal design for cluster randomized trials. *Psychological Methods*, *2*(2), 173–185. https://doi.org/10.1037/1082-989X.2.2.173

Raudenbush, S. W., & Liu, X. (2000). Statistical power and optimal design for multisite randomized trials. *Psychological Methods*, *5*(2), 199–213. https://doi.org/10.1037/1082-989x.5.2.199

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

Raudenbush, S. W., Martínez, A., & Spybrook, J. (2007). Strategies for improving precision in group-randomized experiments. *Educational Evaluation and Policy Analysis*, *29*(1), 5–29. https://doi.org/10.3102/0162373707299460

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

Shen, Z., Kelcey, B. (2020a). *odr: Optimal design and statistical power of multilevel randomized trials (Version 1.0.2) [Software]*. https://cran.r-project.org/package=odr.

Shen, Z., & Kelcey, B. (2020b). Optimal sample allocation in multisite randomized trials. *The Journal of Experimental Education*, 1–19. https://doi.org/https://doi.org/10.1080/00220973.2020.1830361

Shen, Z., & Kelcey, B. (2020c). Optimal sample allocation under unequal costs in cluster-randomized trials. *Journal of Educational and Behavioral Statistics*, *45*(4), 446–474. https://doi.org/10.3102/1076998620912418

Westine, C. D. (2016). Finding efficiency in the design of large multisite evaluations: Estimating variances for science achievement studies. *American Journal of Evaluation*, *37*(3), 311–325. https://doi.org/10.1177/1098214015624014

Spybrook, J., & Raudenbush, S. W. (2009). An examination of the precision and technical accuracy of the first wave of group-randomized trials funded by the Institute of Education Sciences. *Educational Evaluation and Policy Analysis*, *31*(3), 298–318. https://doi.org/10.3102/0162373709339524

Spybrook, J., Shi, R., & Kelcey, B. (2016). Progress in the past decade: An examination of the precision of cluster randomized trials funded by the US Institute of Education Sciences. *International Journal of Research & Method in Education*, *39*(3), 255–267. https://doi.org/10.1080/1743727X.2016.1150454

Zhu, P., Jacob, R., Bloom, H., & Xu, Z. (2012). Designing and analyzing studies that randomize schools to estimate intervention effects on student academic outcomes without classroom-level information. *Educational Evaluation and Policy Analysis*, *34*(1), 45–68. https://doi.org/10.3102/0162373711423786

Appendix

1. The Optimal p Equation

$$\left[nJ\omega(1-R_{3m}^2)p(1-p)+n\rho_2(1-R_2^2)+(1-\rho_2-\rho_3)(1-R_1^2)\right]\left[J\left(c_1^Tn+c_2^T\right)-J(c_1n+c_2)\right]$$
$$p(1-p)-(1-2p)\left[(1-p)J(c_1n+c_2)+pJ\left(c_1^Tn+c_2^T\right)+c_3\right]\left[n\rho_2(1-R_2^2)+(1-\rho_2-\rho_3)(1-R_1^2)\right]=0 \tag{A1}$$

2. Derivation of the Optimal p Expression

Letting the first-order partial derivatives of σ_d^2 in Equation 21 with respect to p be 0, we have

$$\begin{aligned}
\frac{d}{dp}\sigma_\delta^2 &= \frac{d}{dp}\left[\frac{nJ\omega\left(1-R_\omega^2\right)p(1-p)+n\rho_2\left(1-R_2^2\right)+(1-\rho_2-\rho_3)\left(1-R_1^2\right)}{p(1-p)nJK}\right.\\
&\quad\left.\times\frac{(1-p)J(c_1n+c_2)+pJ\left(c_1^Tn+c_2^T\right)+c_3}{m}\right]\\
&=\frac{nJ\omega\left(1-R_\omega^2\right)p(1-p)+n\rho_2\left(1-R_2^2\right)+(1-\rho_2-\rho_3)\left(1-R_1^2\right)}{p(1-p)nJK}\\
&\quad\times\frac{d}{dp}\frac{(1-p)J(c_1n+c_2)+pJ\left(c_1^Tn+c_2^T\right)+c_3}{m}\\
&\quad+\frac{(1-p)J(c_1n+c_2)+pJ\left(c_1^Tn+c_2^T\right)+c_3}{m}\\
&\quad\times\frac{d}{dp}\frac{nJ\omega\left(1-R_\omega^2\right)p(1-p)+n\rho_2\left(1-R_2^2\right)+(1-\rho_2-\rho_3)\left(1-R_1^2\right)}{p(1-p)nJK}\\
&=\frac{1}{m}\left\{\frac{nJ\omega\left(1-R_\omega^2\right)p(1-p)+n\rho_2\left(1-R_2^2\right)+(1-\rho_2-\rho_3)\left(1-R_1^2\right)}{p(1-p)nJK}\right.\\
&\quad\times\left[J\left(c_1^Tn+c_2^T\right)-J(c_1n+c_2)\right]+\left[(1-p)J(c_1n+c_2)+pJ\left(c_1^Tn+c_2^T\right)+c_3\right]\\
&\quad\times\frac{1}{nJK}\times\frac{nJ\omega\left(1-R_\omega^2\right)(1-2p)p(1-p)-(1-2p)}{p^2(1-p)^2}\\
&\quad\left.\times\left[nJ\omega\left(1-R_\omega^2\right)p(1-p)+n\rho_2\left(1-R_2^2\right)+(1-\rho_2-\rho_3)\left(1-R_1^2\right)\right]\right\}=0
\end{aligned}$$

Multiplying each side by $p^2(1-p)^2\frac{m}{nJK}$, we have

$$\begin{aligned}
&\left[nJ\omega\left(1-R_\omega^2\right)p(1-p)+n\rho_2\left(1-R_2^2\right)\right.\\
&\quad\left.+(1-\rho_2-\rho_3)\left(1-R_1^2\right)\right]\left[J\left(c_1^Tn+c_2^T\right)-J(c_1n+c_2)\right]p(1-p)\\
&\quad+\left[(1-p)J(c_1n+c_2)+pJ\left(c_1^Tn+c_2^T\right)+c_3\right]\left\{nJ\omega\left(1-R_\omega^2\right)(1-2p)p(1-p)\right.\\
&\quad\left.-(1-2p)\left[nJ\omega\left(1-R_\omega^2\right)p(1-p)+n\rho_2\left(1-R_2^2\right)+(1-\rho_2-\rho_3)\left(1-R_1^2\right)\right]\right\}\\
&\quad=0
\end{aligned}$$

$$\begin{aligned}&\left[nJ\omega\left(1-R_{\omega}^{2}\right)p(1-p)+n\rho_{2}\left(1-R_{2}^{2}\right)+(1-\rho_{2}-\rho_{3})\left(1-R_{1}^{2}\right)\right]\left[J\left(c_{1}^{T}n+c_{2}^{T}\right)\right.\\&\quad\left.-J(c_{1}n+c_{2})\right]p(1-p)\\&\quad+\left[(1-p)J(c_{1}n+c_{2})+pJ\left(c_{1}^{T}n+c_{2}^{T}\right)+c_{3}\right]\left\{nJ\omega\left(1-R_{\omega}^{2}\right)(1-2p)p(1-p)\right.\\&\quad\left.-(1-2p)\left[nJ\omega\left(1-R_{\omega}^{2}\right)p(1-p)+n\rho_{2}\left(1-R_{2}^{2}\right)+(1-\rho_{2}-\rho_{3})\left(1-R_{1}^{2}\right)\right]\right\}\\&\quad=0\end{aligned}$$

Further combining like terms, we have Equation A1.

3. Derivation of the Optimal n Expression

Letting the first-order partial derivatives of σ_{d}^{2} in Equation 21 with respect to n be 0, we have

$$\begin{aligned}\frac{d}{dn}\sigma_{\delta}^{2}&=\frac{d}{dn}\left[\frac{nJ\omega\left(1-R_{\omega}^{2}\right)p(1-p)+n\rho_{2}\left(1-R_{2}^{2}\right)+(1-\rho_{2}-\rho_{3})\left(1-R_{1}^{2}\right)}{p(1-p)nJ}\right.\\&\qquad\left.\times\frac{(1-p)J(c_{1}n+c_{2})+pJ\left(c_{1}^{T}n+c_{2}^{T}\right)+c_{3}}{m}\right]\\&=\frac{1}{p(1-p)Jm}\frac{d}{dn}\left\{\frac{nJ\omega\left(1-R_{\omega}^{2}\right)p(1-p)+n\rho_{2}\left(1-R_{2}^{2}\right)+(1-\rho_{2}-\rho_{3})\left(1-R_{1}^{2}\right)}{n}\right.\\&\quad\left.\times\left[(1-p)J(c_{1}n+c_{2})+pJ\left(c_{1}^{T}n+c_{2}^{T}\right)+c_{3}\right]\right\}\\&=\frac{1}{p(1-p)Jm}\left\{\left[(1-p)J(c_{1}n+c_{2})+pJ\left(c_{1}^{T}n+c_{2}^{T}\right)+c_{3}\right]\right.\\&\quad\times\frac{d}{dn}\frac{nJ\omega\left(1-R_{\omega}^{2}\right)p(1-p)+n\rho_{2}\left(1-R_{2}^{2}\right)+(1-\rho_{2}-\rho_{3})\left(1-R_{1}^{2}\right)}{n}\\&\quad+\frac{nJ\omega\left(1-R_{\omega}^{2}\right)p(1-p)+n\rho_{2}\left(1-R_{2}^{2}\right)+(1-\rho_{2}-\rho_{3})\left(1-R_{1}^{2}\right)}{n}\\&\quad\left.\times\frac{d}{dn}\left[(1-p)J(c_{1}n+c_{2})+pJ\left(c_{1}^{T}n+c_{2}^{T}\right)+c_{3}\right]\right\}\\&=\frac{1}{p(1-p)Jm}\left\{\left[(1-p)J(c_{1}n+c_{2})+pJ\left(c_{1}^{T}n+c_{2}^{T}\right)+c_{3}\right]\right.\\&\quad\times\left\{\frac{\left[J\omega\left(1-R_{\omega}^{2}\right)p(1-p)+\rho_{2}\left(1-R_{2}^{2}\right)\right]n}{n^{2}}\right.\\&\quad\left.-\frac{\left[nJ\omega\left(1-R_{\omega}^{2}\right)p(1-p)+n\rho_{2}\left(1-R_{2}^{2}\right)+(1-\rho_{2}-\rho_{3})\left(1-R_{1}^{2}\right)\right]}{n^{2}}\right\}\\&\quad\left.+\frac{nJ\omega\left(1-R_{\omega}^{2}\right)p(1-p)+n\rho_{2}\left(1-R_{2}^{2}\right)+(1-\rho_{2}-\rho_{3})\left(1-R_{1}^{2}\right)}{n}\times\left[(1-p)Jc_{1}+pJc_{1}^{T}\right]\right\}\\&=0\end{aligned}$$

Multiplying each side by $n^{2}p(1-p)Jm$, we have

$$\begin{aligned}&\left[(1-p)J(c_{1}n+c_{2})+pJ\left(c_{1}^{T}n+c_{2}^{T}\right)+c_{3}\right]\left\{\left[J\omega\left(1-R_{\omega}^{2}\right)p(1-p)+\rho_{2}\left(1-R_{2}^{2}\right)\right]n\right.\\&\quad\left.-\left[nJ\omega\left(1-R_{\omega}^{2}\right)p(1-p)+n\rho_{2}\left(1-R_{2}^{2}\right)+(1-\rho_{2}-\rho_{3})\left(1-R_{1}^{2}\right)\right]\right\}\\&\quad+\left[nJ\omega\left(1-R_{\omega}^{2}\right)p(1-p)+n\rho_{2}\left(1-R_{2}^{2}\right)+(1-\rho_{2}-\rho_{3})\left(1-R_{1}^{2}\right)\right]\left[(1-p)Jc_{1}\right.\\&\quad\left.+pJc_{1}^{T}\right]n=0\end{aligned}$$

Further combining like terms, we have

$$\left[J\omega\left(1-R_{\omega}^{2}\right)p(1-p)+\rho_{2}\left(1-R_{2}^{2}\right)\right]\left[(1-p)Jc_{1}+pJc_{1}^{T}\right]n^{2} - \left[(1-p)Jc_{2}+pJc_{2}^{T}+c_{3}\right](1-\rho_{2}-\rho_{3})\left(1-R_{1}^{2}\right)=0$$

Further solving for n, we have Equation 22.

4. Derivation of the Optimal J Expression

Letting the first-order partial derivatives of σ_d^2 in Equation 21 with respect to J be 0, we have

$$\begin{aligned}
\frac{d}{dJ}\sigma_{\delta}^{2} &= \frac{d}{dJ}\left[\frac{nJ\omega\left(1-R_{\omega}^{2}\right)p(1-p)+n\rho_{2}\left(1-R_{2}^{2}\right)+(1-\rho_{2}-\rho_{3})\left(1-R_{1}^{2}\right)}{p(1-p)nJ}\right. \\
&\qquad \left.\times\frac{(1-p)J(c_{1}n+c_{2})+pJ\left(c_{1}^{T}n+c_{2}^{T}\right)+c_{3}}{m}\right] \\
&= \frac{1}{p(1-p)nm}\frac{d}{dJ}\left\{\frac{nJ\omega\left(1-R_{\omega}^{2}\right)p(1-p)+n\rho_{2}\left(1-R_{2}^{2}\right)+(1-\rho_{2}-\rho_{3})\left(1-R_{1}^{2}\right)}{J}\right. \\
&\qquad \left.\times\left[(1-p)J(c_{1}n+c_{2})+pJ\left(c_{1}^{T}n+c_{2}^{T}\right)+c_{3}\right]\right\} \\
&= \frac{1}{p(1-p)nm}\left\{\left[(1-p)J(c_{1}n+c_{2})+pJ\left(c_{1}^{T}n+c_{2}^{T}\right)+c_{3}\right]\right. \\
&\qquad \times\frac{d}{\mathrm{d}J}\frac{nJ\omega\left(1-R_{\omega}^{2}\right)p(1-p)+n\rho_{2}\left(1-R_{2}^{2}\right)+(1-\rho_{2}-\rho_{3})\left(1-R_{1}^{2}\right)}{J} \\
&\qquad +\frac{nJ\omega\left(1-R_{\omega}^{2}\right)p(1-p)+n\rho_{2}\left(1-R_{2}^{2}\right)+(1-\rho_{2}-\rho_{3})\left(1-R_{1}^{2}\right)}{J} \\
&\qquad \left.\times\frac{d}{dJ}\left[(1-p)J(c_{1}n+c_{2})+pJ\left(c_{1}^{T}n+c_{2}^{T}\right)+c_{3}\right]\right\} \\
&= \frac{1}{p(1-p)nm}\left\{\left[(1-p)J(c_{1}n+c_{2})+pJ\left(c_{1}^{T}n+c_{2}^{T}\right)+c_{3}\right]\right. \\
&\qquad \times\frac{n\omega\left(1-R_{\omega}^{2}\right)p(1-p)J-\left[nJ\omega\left(1-R_{\omega}^{2}\right)p(1-p)+n\rho_{2}\left(1-R_{2}^{2}\right)+(1-\rho_{2}-\rho_{3})\left(1-R_{1}^{2}\right)\right]}{J^{2}} \\
&\qquad +\frac{nJ\omega\left(1-R_{\omega}^{2}\right)p(1-p)+n\rho_{2}\left(1-R_{2}^{2}\right)+(1-\rho_{2}-\rho_{3})\left(1-R_{1}^{2}\right)}{J}\times\left[(1-p)(c_{1}n+c_{2})\right. \\
&\qquad \left.\left.+p\left(c_{1}^{T}n+c_{2}^{T}\right)\right]\right\}=0
\end{aligned}$$

Multiplying each side by $J^2p(1-p)nm$, we have

$$\begin{aligned}
&\left[(1-p)J(c_{1}n+c_{2})+pJ\left(c_{1}^{T}n+c_{2}^{T}\right)+c_{3}\right]\left\{n\omega\left(1-R_{\omega}^{2}\right)p(1-p)J\right. \\
&\quad \left.-\left[nJ\omega\left(1-R_{\omega}^{2}\right)p(1-p)+n\rho_{2}\left(1-R_{2}^{2}\right)+(1-\rho_{2}-\rho_{3})\left(1-R_{1}^{2}\right)\right]\right\} \\
&\quad +\left[nJ\omega\left(1-R_{\omega}^{2}\right)p(1-p)+n\rho_{2}\left(1-R_{2}^{2}\right)\right. \\
&\quad \left.+(1-\rho_{2}-\rho_{3})\left(1-R_{1}^{2}\right)\right]\left[(1-p)(c_{1}n+c_{2})+p\left(c_{1}^{T}n+c_{2}^{T}\right)\right]J=0
\end{aligned}$$

Further combining like terms, we have

$$\left[n\omega\left(1-R_{\omega}^{2}\right)p(1-p)\right]\times\left[(1-p)(c_{1}n+c_{2})+p\left(c_{1}^{T}n+c_{2}^{T}\right)\right]J^{2} - c_{3}\left[n\rho_{2}\left(1-R_{2}^{2}\right)+(1-\rho_{2}-\rho_{3})\left(1-R_{1}^{2}\right)\right]=0$$

Further solving for J, we have Equation 23.