THE JOURNAL OF EXPERIMENTAL EDUCATION
https://doi.org/10.1080/00220973.2018.1533796

Strategies for Efficient Experimental Design in Studies Probing 2-1-1 Mediation

Ben Kelcey and Zuchao Shen

University of Cincinnati, Cincinnati, OH, USA

ABSTRACT

When well-implemented, mediation analyses play a critical role in probing theories of action because their results help lay the ground work for the critical development of a treatment and the iterative advancement of theories that are foundational to a discipline. Despite strong interest in designs that incorporate mediation, few studies have developed effective and efficient strategies to plan experiments examining multilevel mediation. We probe several design strategies for cluster-randomized designs and derive sampling plans that maximize power under cost constraints. The results suggest that among the more durable design strategies for mediation is covariance adjustment on variables predictive of the outcome and optimal sample allocation. The statistical power and optimal sample allocation results are implemented in the *R* package *PowerUpR*.

KEYWORDS

Indirect effects; mediation; multilevel models; optimal design; power; sample-size determination

STUDIES THAT INVESTIGATE the mechanisms or intermediate actions through which the influence of a treatment operates on an outcome provide evidence as to how and why a treatment comes to prove (in)effective. The type of evidence supplied by mediation analyses often lays the groundwork for critical development of a specific treatment but also for the iterative development and scientific advancement of theories that are foundational to a discipline (e.g., Krull & MacKinnon, 1999). For example, (multilevel) mediation studies have been employed to investigate mental health (Kozlowski & Klein, 2000), steroid use (Krull & MacKinnon, 1999), implementation of evidence-based practice (Aarons, Hurlburt, & Horwitz, 2011), teachers' practice (Kersting, Givvin, Thompson, Santagata, & Stigler, 2012), teacher knowledge and instruction (Kelcey, Hill, & Chin, in press), and academic achievement (Curenton, Dong, & Shen, 2015). Within the context of education and many other social disciplines, mediation analyses often take on a multilevel composition because of the multilevel social structure inherent in many settings (e.g., Krull & MacKinnon, 1999; Kelcey, Dong, Spybrook & Shen, 2017).

Despite widespread interest and development in designs that facilitate mediation analysis, there has been minimal literature outlining the effective and efficient a priori design of studies to detect mediation effects in cluster-randomized designs (Institute of Education Sciences [IES], 2013). A major gap in the literature in terms of planning multilevel mediation studies has been the lack of formulas and statistical approaches to identify designs and strategies that are judicious in their use of resources yet well positioned to generate decisive evidence concerning the theoretical mechanisms of a treatment. For instance, a key component to ensuring that multilevel mediation studies are well designed is identifying sampling schemes—the total sample size and the optimal

CONTACT Ben Kelcey ben.kelcey@gmail.com 6380 Teacher-Dyer, 2160 McMicken Cir, Cincinnati, OH 45221, USA.

allocation of that total sample across levels—that carefully use resources to achieve a sufficient level of power to detect the sequence of relationships used to trace the causal pathways from the treatment through the mediator to the outcome.

In this study, we probe the statistical power formulas to gain conceptual insight into the functional relationships they have with key governing parameters for cluster-randomized trials under three tests. We then derive sample-allocation formulas that optimize the power of each test to detect mediation effects. The structure of our report is divided into four primary sections followed by a discussion. In the first section we introduce a working example and outline the path models used to estimate multilevel mediation effects and provide a summary of the primary assumptions that support inferences. In the second section, we establish statistical power and optimal design formulas for the overall mediation effect for each test. In the third and fourth sections, we extend the results to consider the lower- and upper-level subcomponents of the overall mediation effects when researchers wish to adopt additional assumptions and conceptually decompose the overall mediation effect. We finish with an example and discussion.

Framework

Before outlining the statistical models, we ground our work in a substantive example. We consider the design of a multilevel mediation study that assesses the potential pathways through which participation in a school-based program (treatment) intended to improve student achievement (outcome) operates by improving student behavior (mediator). For an example found in the literature, consider the Raising Healthy Children (RHC) program (Fleming, Harachi, Catalano, Haggerty, & Abbott, 2001). One key component of this program is services that proactively "address developmentally salient risk and protective factors that are precursors to problem behavior" (Fleming et al., 2001, p. 657). More generally, one component of the theory of action underlying the program is that the scaffolded development of prosocial behavior through education and support services reduces problem behavior, develops a sense of belonging, and creates a more productive academic environment. Let us consider a school-randomized design that nests students within schools and assigns schools at random to participate in the RHC program or a control condition. In turn, consider a series of explanatory questions that seek to examine the degree to which the impact of participation in the RHC program (school-level treatment) on student achievement (individual-level outcome) operates through student behavior (individual-level mediator).

In this example, student behavior represents an individual-level mediator because it describes differences among students. However, the collection of behaviors adopted by students within a school can serve to further differentiate schools by faciliating peer effects. For example, prior research has demonstrated that schools that collectively maintain more-positive behavioral environments promote more effective schooling (Wagner & Ruch, 2015). More generally, many theories underlying programs in education specifically draw on and leverage positive environments to affect student outcomes (e.g., Kozlowski & Klein, 2000).

The prevailing approach in characterizing the collective student behavior at a school is to summarize student behaviors using the school average (e.g., Raudenbush & Bryk, 2002). For instance, in our multilevel mediation example, we might consider how participation in the RHC program influences student achievement by examining how participation produces changes in individual student behaviors and how participation modifies the larger school context generated by the collective changes in student behavior. For this reason, many studies considering individual-level mediators assess how school-level treatments affect student-level outcomes by operating through both changes in individual student attitudes and the collection of attitudes at a school

(e.g., Nagengast & Marsh, 2012; Kunst, Fischer, Sidanius, & Thomsen, 2017; Saarento, Boulton, & Salmivalli, 2015; Stabler, Dumont, Becker, & Baumert, 2017).

Models

We describe the models associated with multilevel mediation using our RHC program example. We begin with a multilevel formulation that employs a system of linear mixed models (Pituch & Stapleton, 2012; Raudenbush & Bryk, 2002; VanderWeele, 2010; Zhang, Zyphur, & Preacher, 2009). Our initial model draws on centering student-level variables within school (or group-mean centering) because this is the most common approach in the literature and is useful for disentangling mediation effects across levels (e.g., Zhang, Zyphur, & Preacher, 2009; Pituch & Stapleton, 2012). However, as we subsequently outline, using raw values or centering on the grand mean across schools yields equal parameter estimates under these formulations (Kreft, de Leeuw, & Aiken, 1995). In the context of a student-level mediator, the mediator model is

$$\begin{aligned} M_{ij} &= \pi_{0j} + \pi_1\left(X_{ij} - \bar{X}_j\right) + \pi_2 V_{ij} + \varepsilon_{ij}^M \qquad \varepsilon_{ij}^M \sim N\left(0, \sigma_{M|}^2\right). \\ \pi_{0j} &= \zeta_{00} + aT_j + \zeta_{01} W_j + \zeta_{02}\bar{X}_j + u_{0j}^M \qquad u_{0j}^M \sim N\left(0, \tau_{M|}^2\right). \end{aligned} \tag{1}$$

M_{ij} represents the mediator value for student i in school j, X_{ij} is a student-level covariate that potentially varies across students and schools (with π_1 as its path coefficient), V_{ij} is a student-level covariate that varies only across students (i.e., no school-level variation) with coefficient π_2, and $\bar{X}_j$ is the school-level variable or mean aggregate of the student-level variable (with ζ_{02} as its path coefficient), T_j as the treatment assignment coded as $\pm 1/2$ with associated path coefficient a, W_j as a school-level covariate with ζ_{01} as its path coefficient, ε_{ij}^M as the error term, and u_{0j}^M as the school-specific random effects. Applied to our running example, a maps out how participation in the RHC program produces changes in student behavior.

The outcome model parallels the previous model such that,

$$\begin{aligned} Y_{ij} &= \beta_{0j} + b_1\left(M_{ij} - \bar{M}_j\right) + \beta_1\left(X_{ij} - \bar{X}_j\right) + \beta_2 V_{ij} + \varepsilon_{ij}^Y \qquad \varepsilon_{ij}^Y \sim N\left(0, \sigma_{Y|}^2\right). \\ \beta_{0j} &= \gamma_{00} + B\bar{M}_j + c'T_j + \gamma_{01} W_j + \gamma_{02}\bar{X}_j + u_{0j}^Y \qquad u_{0j}^Y \sim N\left(0, \tau_{Y|}^2\right). \end{aligned} \tag{2}$$

We use Y_{ij} as the outcome for student i in school j, $M_{ij} - \bar{M}_j$ as the school-centered student-level mediator with coefficient b_1, $\bar{M}_j$ as the mean of the mediator in school j with path coefficient B, c' as the treatment-outcome conditional path coefficient, and u_{0j}^Y and ε_{ij}^Y as the Level 2 and Level 1 error terms. Returning to our example, the B path describes how changes in student behavior are (conditionally) correlated with students' achievement (see below for additional discussion).

Assumptions

Literature probing the interpretation of indirect effects resulting from the aforementioned models as evidence for mediation has been careful to detail key assumptions required to draw inferences regarding indirect effects. In addition to the usual model-based assumptions (e.g., correct specification, linearity, homogeneity), we outline three primary assumptions below and refer to literature for more detailed descriptions (e.g., VanderWeele, 2010).

A first core assumption supporting interpretation is sequential ignorability. Although random assignment of the treatment balances (un)observed confounding variables between groups for the treatment-mediator and treatment-outcome paths, it does not directly balance those variables that may confound the mediator-outcome path. Inferences concerning mediation depend on the

validity of the sequential ignorability assumption that outlines the conditional independence of the potential mediator and outcome responses given observed variables (VanderWeele, 2010). Put differently, the sequential ignorability requirement suggests that once we have controlled for the observed confounding (e.g., from X), there are no unobserved variables that further confound the paths. Returning to our example, if we had randomly assigned schools to treatment conditions but our theory suggested that the type of school (e.g., public vs. private) influences student behavior and student achievement, then sequential ignorability (and unbiased estimates of mediation effects) would be obtainable only if we conditioned on school type in our model.

A second important assumption in this setting is the stable unit treatment value assumption (SUTVA). There are two core components in this assumption. The first is that there exists only one version of the treatment. In our example, this assumption suggests that there is minimal variation in the delivery and content of the RHC program across schools.

The second component of SUTVA is that a student's potential outcome and mediator values are not contingent upon the treatment condition and mediator values of other students (VanderWeele, 2010). Specifically, the consideration of a student-level mediator requires the adoption of an individual-level version of SUTVA because mediator values vary within schools (VanderWeele, 2010). Identification of mediation effects in the 2-1-1 mediation case is ostensibly predicated on a student's potential outcomes not being contingent on schoolmates' mediator values or the treatment assignment of students at other schools (i.e., no interference).

To examine this limitation, literature has considered the pathways through which the mediator values of schoolmates are likely to sway the potential outcomes of students (e.g., Hong & Raudenbush, 2006; VanderWeele, 2010). A simplification arises when we can theoretically limit student's influence to only their schoolmates (those at the same school). Such a scenario arises when, for example, schools operate with a reasonable level of independence or a minimal level of interaction with other schools. When this assumption holds, a student's potential outcomes (e.g., achievement) depend only on the mediator values of schoolmates.

A second simplification arises when we can reasonably identify a scalar function through which the sway of schoolmates on a student is likely to operate. The most common operationalization of this influence is through a school-level mean of the student mediator values in a school (Raudenbush & Bryk, 2002; Pituch & Stapleton, 2012; VanderWeele, 2010). Applied to our example, this might suggest that the influence of schoolmates' behaviors on a student's achievement arises primarily through the collective or average behavior at that school.

For instance, schoolmates' behavior may positively influence a student's achievement. Such an example might suggest that exposing a school to the program changes individual student behaviors and the collective behaviors of students and that, subsequently, those individual and collective changes in behavior contribute to a student's achievement level. If those changes in the environment can be adequately captured by the average student behavior, this assumption is satisfied. Alternatively, if the changes in environment were influenced by, for example, the variability of behavior in a school in addition to the average student behavior, then this assumption would be satisfied only if our models included the school average behavior and the dispersion of that behavior within a school.

A third assumption is that there is no mediator-by-treatment interaction. Contemporary mediation analysis has been careful to consider the possibility that the treatment impacts the magnitude of the mediator-outcome relationship (e.g., magnitude of the B path) in addition to the mediator. In our applied example, for instance, this type of moderation would exist if participation in the RHC program intensified the link between student behavior (mediator) and achievement (outcome) as well as improving behavior. This assumption is relaxed by introducing a treatment-by-mediator interaction in the outcome model. In our subsequent analyses, we simplify the presentation by omitting the interaction term; however, including it is a simple extension (e.g., Kelcey, Dong, Spybrook, & Cox, 2017; Kelcey, Dong, Spybrook, & Shen, 2017).

Mediation effect

In the literature, there are two prevailing perspectives on expressing the movement of effects from a treatment to an outcome as they operate via an intermediate individual-level variable when using cluster-level assignment. In a first perspective, mediational analyses with a mediator measured at the individual-level consider how a treatment operates through both the original individual-level mediator and its cluster-level aggregate (e.g., Pituch & Stapleton, 2012; VanderWeele, 2010; Krull & MacKinnon, 2001; Talloen et al., 2016; Kelcey, Dong, & Spybrook, 2018). More specifically, this perspective examines how a treatment generates changes in an outcome by working through the individual-level mediator values (i.e., differences among students within a school) and in the school-level aggregate of the mediator (i.e., differences among schools). Consequently, this perspective studies the prospect that the treatment works differently at the student level and the school level (e.g., differing degrees or directions). Such a perspective considers the possibility that the (overall) mediation effect can be at least conceptually separated into a component resulting from the collective changes in the mediator for all students in a school and a component resulting from changes in the mediator values of specific students.

In a second perspective, researchers consider only the overlapping covariances at the school level because a school-level treatment can covary only with the school-level variation in the mediator (e.g., Zhang, Zyphur, & Preacher, 2009; Kelcey, Dong, Spybrook, & Cox, 2017). The result is that the treatment-mediator covariance can covary only with the school-level outcome variance. That is, because students in the same school are assigned identical treatment conditions, decomposing the degree to which the treatment operates through the student level versus school level is not identified and opposes the statistical theory underlying such mediation.

Ostensibly these perspectives are at odds with each other because the former privileges a substantive perspective whereas the latter privileges a statistical perspective (Kelcey, Dong, & Spybrook, 2018). From a substantive standpoint, prior literature considers it natural and useful to descriptively partition such relationships (e.g., Pituch & Stapleton, 2012). From a statistical standpoint, literature has suggested that the decomposition incurs additional assumptions and is statistically incoherent (e.g., Zhang, Zyphur, & Preacher, 2009).

From a pragmatic perspective, the practical difference between the methods reduces, whether one wishes to draw on additional assumptions, to separate out the (overall) mediation effect into contextual and individual components or not. We provide additional discussion regarding these disparities below; however, from a practical viewpoint, both perspectives privilege, consider, and evaluate the overall mediation effect but the decomposition perspective additionally delineates the overall mediation effect into a unique school-level component (contextual) and a unique student-level component. Our investigation in this study focuses primarily on the overall mediation effect but also presents derivations (but limited discussion) for the separated effects. The net implications for our analysis are minimal because the underlying derivations for statistical power and optimal sampling are identical for the two perspectives: If a researcher adopts the overall mediation-only approach, she or he can use the results developed below regarding the overall mediation effect and ignore the auxiliary results. If instead a researcher is interested in designing a study to specifically track the lower- or upper-level mediation routes, the researcher can draw on the results below that are specific to each path.

Overall mediation effect

Mediation effect

Under the assumption that students influence only their own schoolmates and do so through the collective mean and the other aforementioned assumptions, we can consider the three types of mediation that researchers might examine when an individual-level mediator is of interest. The

first category we focus on is the cumulative or overall mediation effect that is composed of the unique contextual (upper-level) and individual (lower-level) mediation effects (described below). We use the overall mediation effect to describe, for example, how changes in behaviors (individual or collective or both) brought about by a school's participation in the RHC program manifest as changes in student achievement.

Under the centering-within-school approach (as in expression (2)), the coefficient attached to the school mediator mean (B) represents the sum of the mediator's student-level relationship with the outcome (b_1) and the mediator's unique contextual relationship with the outcome (b_2; see below; Kreft, de Leeuw, & Aiken, 1995). That is, the B coefficient captures the total (individual plus contextual) influence of the mediator on the outcome as it operates through the student-level mediator values and the mean scalar function of the school mediator values.

As a result, the product of the a and B coefficients can be used to capture the overall mediation effect of the treatment on the outcome as it operates through the student mediator and the school mean of the mediator values (Zhang, Zyphur, & Preacher, 2009; Pituch & Stapleton, 2012). Given the centering-within-schools parameterization, prior literature (VanderWeele, 2010, VanderWeele & Vansteelandt, 2009; Pituch & Stapleton, 2012) has shown that we can obtain an estimate of the overall mediation effect (OME) as

$$\text{OME} = a(b_1 + b_2) = aB. \tag{3}$$

In our example, this overall mediation effect quantifies the improvement in student achievement that accrues as a result of program-induced changes in both student and schoolmate behavior.

Having outlined the models and concepts, we next apply this framework to the effective and efficient design of multilevel mediation studies. Below, we outline three test statistics to draw inferences regarding mediation under maximum likelihood estimation (Kelcey, Dong, Spybrook, & Cox, 2017). We describe the power with which each test can detect effects, probe the relationships between power and key parameters, and derive the sampling plans that optimize the power of each test.

Sobel test

The Sobel test contrasts the ratio of the estimated mediation effect (aB) and its standard error to an asymptotic standard normal distribution (Sobel, 1982). For the overall mediation effect, the Sobel test statistic can be formed using

$$z_{aB}^{Sobel} = aB/\sqrt{\sigma_{aB}^2}. \tag{4}$$

where σ_{aB}^2 is the error variance associated with the overall mediation effect. This error variance can be estimated using

$$\sigma_{aB}^2 = a^2\left(\frac{\left(\tau_{Y|}^2 + \sigma_{Y|}^2/n_1\right)}{n_2\left(\tau_{M|}^2 + \sigma_{M|}^2/n_1\right)}\right) + B^2\left(\frac{\tau_{M|}^2 + \sigma_{M|}^2/n_1}{P(1-P)n_2}\right). \tag{5}$$

where n_1 is the number of students per school, n_2 is the number of schools sampled, P is the proportion of schools assigned to the treatment; the remaining terms have been defined previously (Kelcey, Dong, Spybrook, & Cox, 2017). If we normalize the outcome and mediator to have unconditional means of zero and variances of one (i.e., $\tau_M^2 + \sigma_M^2 = \rho_M + (1-\rho_M) = 1$ and $\tau_Y^2 + \sigma_Y^2 = \rho_Y + (1-\rho_Y) = 1$), the school-level mediator and outcome variances serve as the respective intraschool correlation coefficients (ρ_M and ρ_Y).

We can further reframe the error variance as a function of the governing parameters. To do so, we replace the conditional variance terms composing the error variance of the overall

mediation effect as functions of the parameters (Kelcey, Dong, Spybrook, & Cox, 2017). The conditional school-level outcome variance can be expressed as

$$\tau^2_{Y|} = \rho_Y\left(1-R^2_{Y^{L2}_{W,\bar{X}}}\right)-P(1-P)(aB+c')^2-\left[B^2\rho_M\left(1-R^2_{M^{L2}_{W,\bar{X}}}\right)+B^2(1-\rho_M)\left(1-R^2_{M^{L1}_{X,V}}\right)/n_1-a^2B^2P(1-P)\right]. \tag{6}$$

where $R^2_{M^{L2}_{W,\bar{X}}}$ and $R^2_{M^{L1}_{X,V}}$ are the proportions of variance explained in the mediator by the covariates. The conditional student-level outcome variance can be estimated using

$$\sigma^2_{Y|} = (1-\rho_Y)\left(1-R^2_{Y^{L1}_{X,v}}-\left(\frac{(1-\rho_M)}{(1-\rho_Y)}b_1^2\left(1-R^2_{M^{L1}_{X,V}}\right)\right)\right). \tag{7}$$

For the conditional mediator variances we have

$$\tau^2_{M|} = \rho_M\left(1-\left(P(1-P)a^2\right)/\rho_M-R^2_{M^{L2}_{W,\bar{X}}}\right) \quad \text{and} \quad \sigma^2_{M|} = (1-\rho_M)\left(1-R^2_{M^{L1}_{X,V}}\right). \tag{8ab}$$

The power to detect an overall mediation effect using the Sobel test is assessed using

$$P\left(z^{Sobel}_{aB} > z_{critical}\right) = 1-\Phi\left(z_{critical}-z^{Sobel}_{aB}\right)+\Phi\left(-z_{critical}-z^{Sobel}_{aB}\right). \tag{9}$$

Relationships between power and parameters

To assess the functional relationships between each of the parameters and power we can study, for example, rates and monotonicity of the relationships using additional algebra. Such algebraic manipulations, however, still do not provide a clear conceptual sense of the relationships because they result in equally complex expressions. For this reason, we outline the nature of relationships between power and key parameters using a prototypical example that characterizes the complexity of each relationship. However, we are careful to note that in many instances, noted below, the nature of these relationships can be sensitive to the specific set of parameter values.

Let us anchor our probe in the working example established above regarding the discovery of the overall mediation effect when targeting a student-level mediator is of interest. Consider the design of a multilevel mediation study that probes the degree to which the RHC program impacts student achievement by improving student behavior (mediator). In this example, we focus on the total or overall mediation effect that captures the extent to which participation in the program generates changes in student achievement by changing student individual behavior and/or their schoolmates' behavior.

Presume that we have collected empirical estimates of the parameter values based on prior literature such that we anticipate that the unconditional school-level variances for the outcome and mediator are both 0.20 (i.e., $\tau^2_M = \tau^2_Y = 0.20$) and the unconditional student-level variances are 0.80 (i.e., $\sigma^2_M = \sigma^2_Y = 0.80$). Assume that of this unconditional variance, about 10% is explained by covariates at each level for each response (i.e.,$R^2_{Y^{L2}_z} = R^2_{M^{L2}_z} = R^2_{Y^{L1}_z} = R^2_{M^{L1}_z} = 0.10$, with z indicating the appropriate covariates for a response based on the models above). Regarding the effect expectations, consider a study designed to detect a treatment-mediator path coefficient as small as $a = 0.45$, an overall mediator-outcome association of $B = 0.35$ with 0.15 due to the student-level association (b_1 =0.15), and a direct effect of the treatment on the outcome of $c' = 0.05$. Further assume that half of the schools will receive the RHC program ($P = 0.50$).

In probing the relationships, we sequentially freed one key parameter while fixing the remaining parameters to the values noted above. In turn, we plot how statistical power changes as a function of each key parameter. We outline the relationships for the following parameters: (a) the a path coefficient, (b) the B (and b_1) path coefficient, (c) the unconditional outcome school-level variance (τ^2_Y), (d) the unconditional school-level mediator variance (τ^2_M), (e) the

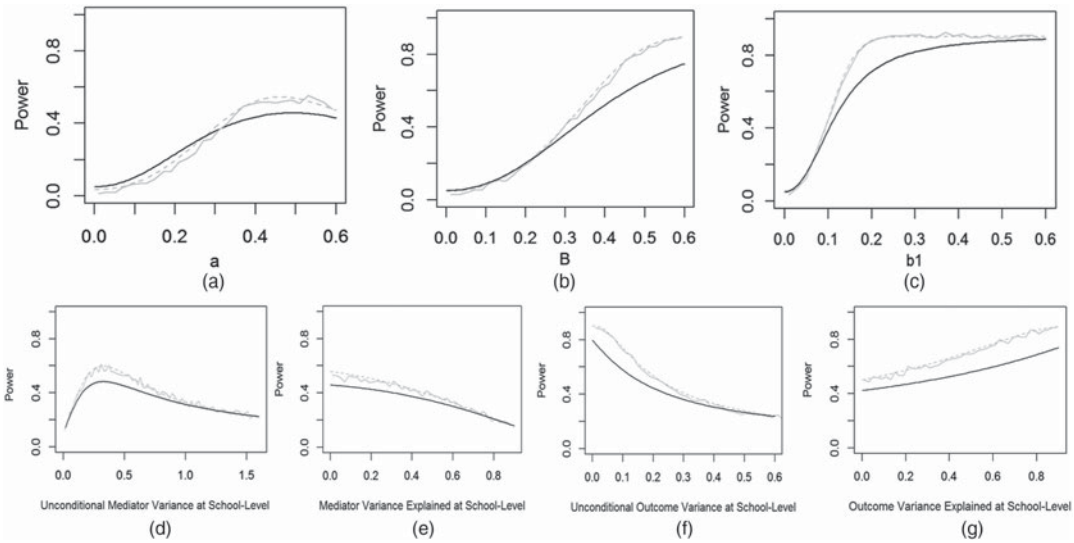

Figure 1. Power under the Sobel test (black) and joint test (gray dashed) as a function of the (a) treatment-mediator path coefficient (a), (b) total mediator-outcome path coefficient (B), (c) student-level mediator-outcome path coefficient (b_1), (d) unconditional mediator variance at the school level (τ^2_M), (e) mediator variance explained at the school level by covariates ($R^2_{M^{L2}_z}$), (f) unconditional outcome variance at the school level (τ^2_Y), and (g) outcome variance explained at the school level by covariates ($R^2_{M^{L2}_z}$).

outcome variance explained by covariates at the school level ($R^2_{Y^{L2}_z}$), (f) the mediator variance explained by covariates at the school level ($R^2_{M^{L2}_z}$), and (g) the number of students per school (n_1). The relationships are outlined in Figure 1 for the Sobel test (and other tests subsequently covered) with a brief summary below.

Treatment-mediator path. The first parameter we examined was the a path coefficient (Figure 1a). For the Sobel test, our example suggests that power is a nonmonotonic function of the a path coefficient. In our example, holding other parameters fixed, the power to detect the overall mediation effect increased when a increased from zero to about 0.5 but then began to decrease after that. This result aligns with past literature on single-level mediation that notes that although increases in the a path produce larger mediation effects, such effect increases are coupled with additional collinearity that inflates the uncertainty of the mediation effect (e.g., Beasley, 2014). In contrast, increases in the magnitude of the mediator-outcome path coefficient produced monotonic gains in power (Figure 1b).

Unconditional mediation variance. We next probed the influence of the unconditional mediator variance at the school level (Figure 1d). The relationship between power and the unconditional mediator variance at the school level in our example was nonmonotonic and was heavily dependent on the values of other parameters. In our example, increases in τ^2_M were associated with increases in power from about zero to about 0.4 (i.e., an intraclass correlation coefficient of about 0.3) with subsequent increases being associated with decreasing power.

Variance explained in the mediator. Similarly, increases in the variance explained in the mediator at the school level by covariates ($R^2_{M^{L2}_z}$) were negatively associated with power (Figure 1e). The nature of these relationships is not immediately intuitive. The complexity arises because the contribution of the unexplained mediator variance to the error variance of the mediation effect is twofold—for the a path, increasing the mediator variance explained decreases $\tau^2_{M|}$ and thus serves to reduce the uncertainty of the mediation effect because it enters into the numerator of the error variance (see Equation 5). In contrast, for the B path, explaining mediator variance with covariates potentially serves to inflate the uncertainty because it enters into the denominator (see Equation 5); however, the impact here depends on how the variance explained in the outcome also changes as a function of those covariates. That is, decreases in σ^2_a attributable to

covariates ($R^2_{M^{L2}_z}$) can be offset by corresponding increases in σ^2_B. However, if increases in $R^2_{M^{L2}_z}$ are paired with increases in $R^2_{Y^{L2}_z}$, the relationship can be almost completely eliminated; increases in $R^2_{M^{L2}_z}$ (and thus $R^2_{Y^{L2}_z}$) would yield virtually no change in power. Overall, the influence of parameters linked to the explained mediator variance varies widely and is heavily dependent on the specific values of other parameters and the ratios of these parameters.

More generally, analysis of the mediator variance explained by covariates suggests that design strategies predicated on covariate selection or "optimizing" this parameter will typically be dubious and highly unlikely to improve inferences for at least two reasons. First and foremost, the sequential ignorability assumption necessitates adjustment for all covariates that confound the mediation relationships. Removing a covariate that confounds the mediation relationships because it may weaken power will simply yield more-precise but biased estimates of the relationships. Second, the results suggest that any potential gains in power obtained by identifying an optimal level of variance explained in the mediator by covariates are highly sensitive to the presumed values of other parameters. As a result, most theoretical gains in efficiency will be quickly subjugated by any misspecification of the remaining parameter values.

Outcome variance. The relationship between power and the unconditional (and conditional) outcome variance was much simpler. With similar constraints, we saw that power was a steadily decreasing function of the unconditional outcome variance at the school level (Figure 1f): The higher the degree of correlation among students in the same school, the lower the power. Similarly, increases in the outcome variance explained at the school level by covariates ($R^2_{Y^{L2}_z}$) tend to produce more-powerful designs (Figure 1g). That is, we can partially reduce the impact of clustering in the outcome by introducing covariates that explain school-level variation. Similar to designs for main effects, the results suggest that one simple design strategy for improving mediation power is to incorporate covariates that explain variation in the outcome (e.g., Kelcey, Phelps, Spybrook, Jones & Zhang, 2017).

Mediator-outcome path. Last, we observed similarly simple and positive relationships between power and the mediator-outcome path coefficients. Increases in the B (and b_1) path coefficient yield greater power (Figure 1b and 1c). These relationships arise primarily because increases in the mediator-outcome relationship enlarge the mediation effect and reduce uncertainty in that effect (see Equation 5).

Optimal sampling. We next investigated the role of the number of students sampled per school (n_1). A key consideration in the design of studies is the efficient use of resources. In most studies, many of the parameters governing the power of a design to detect effects are not malleable (e.g., Kelcey & Phelps, 2013a). For instance, the correlation among students within the same school on the mediator and outcome are structural features of schooling that are difficult to manipulate in addition to implementing a program (e.g., Kelcey & Phelps, 2013b). In experimental design, researchers often can manipulate sample sizes at each level subject to cost constraints (Raudenbush, 1997; Cox & Kelcey, in press). Increasing the number of schools sampled will typically be associated with larger increases in power when compared to increases in the number of students sampled per school. Adding schools, however, typically incurs larger costs.

One strategy detailed in prior literature is to identify the student and school sampling balance that produces the most power within prespecified budgetary constraints. We draw on prior optimal design frameworks and consider a linear cost formulation (Raudenbush, 1997; Konstantopoulos, 2009; Cox & Kelcey, in press):

$$c = c_2 n_2 + c_1 n_2 n_1. \tag{10}$$

Here, we allow c to be the total funds available to conduct a study with c_1 as the cost of sampling each student and c_2 as the cost of sampling each school. Under this cost structure, we can represent the school sample size as a function of the student sample size and the cost ratio:

$$n_2 = c/(c_2 + c_1 n_1) \tag{11}$$

Having defined the school-level sample size as a function of the student-level sample size and cost structure, we can indirectly maximize statistical power under the Sobel test by minimizing the error variance with respect to n_1. We derive the optimal sample allocation under the assumption that the number of clusters is equal across treatment conditions (i.e., $P = 0.5$). However, relaxing this assumption is straightforward. To identify the optimal number of students per school, we can take the first-order derivative of the error variance in terms of n_1 and set it equal to zero. The derivative yields

$$\frac{\partial \sigma^2_{aB}}{\partial n_1} = \frac{1}{c}(\Omega + \Xi + \Psi - Z + \Theta). \tag{12}$$

with

$$\Omega = \frac{\left(4a^2\left(R^2_{M^{L1}_z}-1\right)\sigma^2_M(c_1n_1+c_2)\left(2aBc'n_1-4B^2\left(\left(R^2_{M^{L1}_z}-1\right)\sigma^2_M+\left(R^2_{M^{L2}_z}-1\right)\tau^2_Mn_1\right)-4b_1^2\left(R^2_{M^{L1}_z}-1\right)\sigma^2_M+c'^2n_1+4R^2_{Y^{L1}_z}\sigma^2_Y+4R^2_{Y^{L2}_z}\tau^2_Yn_1-4\sigma^2_Y-4\tau^2_Yn_1\right)\right)}{\left(n_1\left(n_1\left(a^2+4\left(R^2_{M^{L2}_z}-1\right)\tau^2_M\right)+4\left(R^2_{M^{L1}_z}-1\right)\sigma^2_M\right)^2\right)}$$

$$\Xi = \frac{4a^2(c_1n_1+c_2)\left(B^2\left(R^2_{M^{L1}_z}-1\right)\sigma^2_M+b_1^2\left(R^2_{M^{L1}_z}-1\right)\sigma^2_M-R^2_{Y^{L1}_z}\sigma^2_Y+\sigma^2_Y\right)}{n_1\left(n_1\left(a^2+4\left(R^2_{M^{L2}_z}-1\right)\tau^2_M\right)+4\left(R^2_{M^{L1}_z}-1\right)\sigma^2_M\right)}$$

$$\Psi = \frac{\left(a^2c_1\left(2aBc'n_1-4B^2\left(\left(R^2_{M^{L1}_z}-1\right)\sigma^2_M+\left(R^2_{M^{L2}_z}-1\right)\tau^2_Mn_1\right)-4b_1^2\left(R^2_{M^{L1}_z}-1\right)\sigma^2_M+c'^2n_1+4R^2_{Y^{L1}_z}\sigma^2_Y+4R^2_{Y^{L2}_z}\tau^2_Yn_1-4\sigma^2_Y-4\tau^2_Yn_1\right)\right)}{\left(n_1\left(a^2+4\left(R^2_{M^{L2}_z}-1\right)\tau^2_M\right)+4\left(R^2_{M^{L1}_z}-1\right)\sigma^2_M\right)}$$

$$Z = \frac{B^2c_1\left(n_1\left(a^2+\left(4R^2_{M^{L2}_z}-4\right)\tau^2_M\right)+\left(4R^2_{M^{L1}_z}-4\right)\sigma^2_M\right)}{n_1}$$

$$\Theta = \frac{4B^2\left(R^2_{M^{L1}_z}-1\right)\sigma^2_M(c_1n_1+c_2)}{n_1^2}$$

Although a simple closed-form solution identifying the optimal n_1 is not available, we can identify the optimal number of students by setting this derivative equal to zero and solving it numerically. We implement this derivative and its solution in the *R* package PowerUpR.

As with optimal-design considerations for main effects, the optimal sampling scheme for the overall mediation effect under the Sobel test does not depend on the total funds available (c). In line with prior parameters though, the resulting optimal n_1 is a much more complex function of the remaining parameters. To probe the functional relationships between the optimal sampling plan and the parameters noted above, let us return to our working example and assume that half of the sampled schools will be assigned to the RHC program ($P = 0.50$) and our total budget is $c = 500{,}000$ with each additional student costing $c_1 = 100$ and each additional school adding $c_2 = 10{,}000$.

We plot how the optimal sampling plan changes as a function of each key parameter in Figure 2. For the Sobel test, our example suggests that the optimal student sample size is a steadily decreasing function of the a path coefficient (Figure 2a). That is, larger a coefficients will typically suggest sampling more schools in exchange for fewer students per school. For instance, holding other parameters fixed, the optimal number of students per school is nearly $n_1 = 20$ when $a = 0.10$ and decreases to $n_1 = 5$ when $a = 0.60$.

For the B coefficient and the cost ratio, the results suggested that the optimal number of students per school was a monotonically increasing function. The optimal number of students per school increased at a rate similar to but opposite of that of the treatment-mediator relationship. For example, holding other parameters fixed, the optimal number of students per school is nearly $n_1 = 3$ when $B = 0.10$ but increases to about $n_1 = 13$ when $B = 0.50$.

For the cost ratio, the results also indicated that there was a positive relationship between increasing school costs and the optimal number of students per school. When the cost of sampling a school relative to a student increased, optimal sampling favored more students per school over more schools. However, the slope of this relationship diminishes quickly once the cost ratio exceeds about 50 or so (Figure 2c).

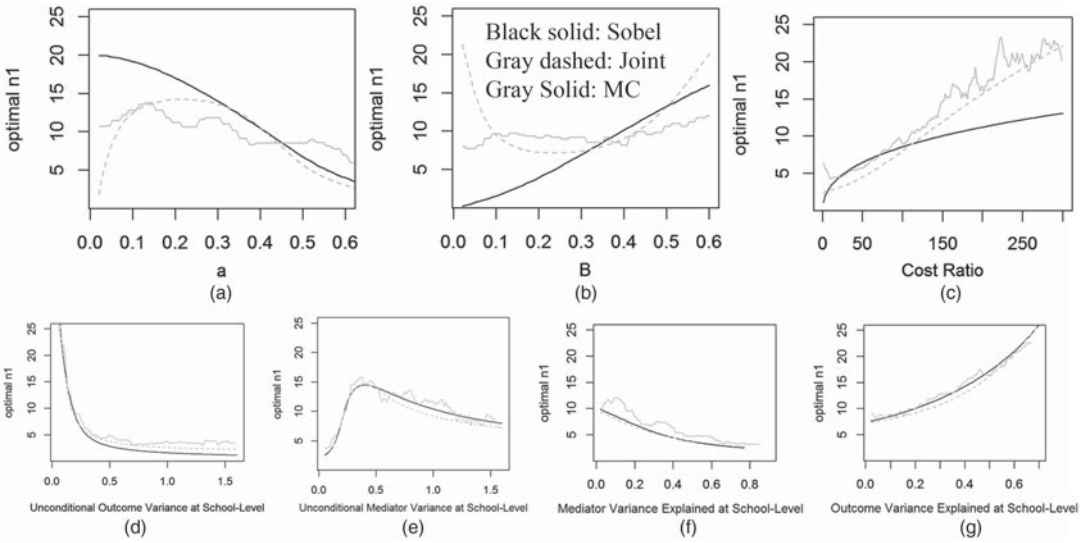

Figure 2. Optimal number of students per school for the overall mediation effect under the Sobel test (black) and joint test (gray dashed) as a function of the (a) treatment-mediator path coefficient (a), (b) total mediator-outcome path coefficient (B), (c) cost ratio (c_2/c_1), (d) unconditional mediator variance (τ^2_M), (e) mediator variance explained by covariates at the school level ($R^2_{M^{L2}_z}$), (f) unconditional outcome variance (τ^2_Y), and (g) outcome variance explained by covariates at the school level ($R^2_{Y^{L2}_z}$).

Last we assessed the influence of the mediator and outcome unconditional and conditional variances. For the outcome, the optimal number of students per school was a monotonically decreasing function of the unconditional outcome school-level variance; when there was substantial clustering within a school, it is more efficient to sample additional schools rather than additional students per school (Figure 2d). Similarly, there was a positive relationship between the optimal number of students per school and the proportion of the outcome variance at the school level explained by covariates. That is, if you can explain the majority of the outcome clustering using covariates ($R^2_{Y^{L2}_z}$), then it is more efficient to give up additional schools to sample additional students in each school (Figure 2g).

Similar to power, the relationship between optimal sampling and mediator variance was more complex (Figure 2ef). From an unconditional school-level mediator variance of zero to about 0.4, an increase in this variance was paired with larger values for the optimal number of students per school. However, subsequent increases in the unconditional mediator variance were coupled with decreases in the optimal number of students per school. Again, although our example characterizes the generally complex nature of this relationship, we note that the specific behavior and inflection points of this relationship are highly dependent on the specific parameter values one considers.

Joint test

Although the asymptotic distribution of the maximum likelihood estimates of the mediation effect is normal, this approximation can be problematic under small sample sizes or with large disparities in the magnitudes of the a and b path coefficients (e.g., MacKinnon, et al., 2004). The net effect of such imprecision is that the Sobel test often incurs Type 1 error rates below the nominal level and produces underpowered tests.

The joint test circumvents the nonnormality that arises with the Sobel test statistic in finite samples and disparate path magnitudes by conducting path-specific subtests (MacKinnon et al., 2004). The joint test assesses mediation by evaluating the simultaneous significance of the

treatment-mediator and mediator-outcome paths. Past literature has widely supported the efficacy of the joint test because it has consistently turned in performances that are comparable to the most sensitive resampling-based tests (e.g., bootstrap methods; Hayes & Scharkow, 2013).

The joint test can be implemented through two simultaneous subtests that compare the ratio of the respective path and its standard error to a normal distribution (Raudenbush & Bryk, 2002; Kenny & Judd, 2014). The test statistics are

$$z_a = a/\sqrt{\sigma_a^2} \quad \text{and} \quad z_B = B/\sqrt{\sigma_B^2}. \tag{13ab}$$

with σ_a^2 and σ_B^2 as the error variances. The error variances for the a and B paths are

$$\sigma_a^2 = \left(\frac{\tau_{M|}^2 + \sigma_{M|}^2/n_1}{P(1-P)n_2}\right) \text{ and } \sigma_B^2 = \frac{\tau_{Y|}^2 + \sigma_{Y|}^2/n_1}{n_2(\tau_{M|}^2 + \sigma_{M|}^2/n_1)}. \tag{14ab}$$

These error variances can also be expressed as direct operations of the path coefficients and the variance explained by covariates (see above and Kelcey, Dong, Spybrook, & Cox, 2017).

Inferences under the joint test are drawn on the basis of both test statistics rejecting their null hypotheses that their respective paths are zero. To derive the power associated with a two-sided joint test (i.e., both paths concurrently nonzero), we consider the product of the power to detect the treatment-mediator path and the power to detect the mediator-outcome path. The resulting power approximation for the overall mediation effects is then calculated as

$$P\left(|z_a| > z_{critical} \& |z_B| > z_{critical}\right) = \left(1 - z(z_{critical} - z_a) + z(-z_{critical} - z_a)\right) * \left(1 - z(z_{critical} - z_B) + z(-z_{critical} - z_B)\right). \tag{15}$$

with z as the respective cumulative standard normal density function and $z_{critical}$ as the corresponding critical value (e.g., 1.96).

As with the Sobel test, we studied the relationships between each of the parameters and power under the joint test. Using the same prototype and values, the relationships are overlain with those of the Sobel test in Figure 1. Overall, the examples suggested that the joint test was consistently more powerful than the Sobel test. However, the results suggested that the functional relationships between power and each of the parameters under the joint test were very similar to those under the Sobel test.

Optimal sampling

The joint test does not directly draw on or specify the error variance of the mediation effect and as a result we cannot identify the sampling plan that maximizes power by simply minimizing the error variance as we did with the Sobel test. Furthermore, even identifying the sampling plan that maximizes the product of the test statistics (i.e., $z_a z_B$) is insufficient to identify the maximum power because changes in power as a function of sample size are nonlinear. As a result, in contrast to steps taken with the Sobel test, we must identify the optimal sampling plan under the joint test by directly maximizing the power function (Expression 15). The derivative of the power function for the joint test is in the appendix. As with the Sobel test, though there is no closed-form solution, we can quickly identify the optimal number of students by solving the derivative numerically. We implement these optimizations in the *R* package PowerUpR.

As with the Sobel test with regard to optimal sampling, we explored the influence of the parameters under the same parameter values to get a better sense of what drives optimal sampling under the joint test (Figure 2). Varying the a path coefficient produced more-complex results than in the Sobel test. The optimal number of students per school increased when the magnitude of the a path strengthened from zero to about 0.25 and then decreased with subsequent gains (Figure 2a). For the B path coefficient we also saw a departure from past behavior

under the Sobel test. Under the joint test there was a nonmonotonic relationship between the B path coefficient and the optimal number of students per school: from about zero to 0.25 the function was decreasing; whereas above 0.25 the function increased.

The remaining parameters behaved directionally similar to those under the Sobel test. However, the functional form or magnitude of the influence of each parameter often differed compared to that under the Sobel test. For example, although the optimal number of students per school increased under the Sobel and joint tests when the cost ratio grew, the increases were much more pronounced under the joint test for a cost range between 1 and 300 (Figure 2c).

Monte Carlo interval test

An alternative to the Wald-like mediation tests is the resampling-based Monte Carlo interval test (MacKinnon et al., 2004; Preacher & Selig, 2012). The Monte Carlo interval test employs the primary path coefficient estimates and their error variances to simulate draws from the posterior distribution of the mediation effect. With a sufficient number of draws, we can estimate the sampling distribution of the estimated mediation effect and can implement the Monte Carlo interval test assessing whether desired confidence intervals include a null mediation effect. This approach has proven valuable in the literature because it accommodates the asymmetries in the distribution of the estimated mediation effect that arise from, for example, small sample sizes or disparate path magnitudes, while returning a robust performance relative to bootstrapping and other methods. A key advantage of the Monte Carlo interval test is that it can be employed without access to full data or, as in our case, during the design phase when no data are available.

Using the normality of the individual maximum likelihood estimates, we apply the Monte Carlo confidence interval test using a multivariate normal distribution for the estimated path coefficients. Path means are set to their estimated values, error variances are set to the aforementioned path-specific error variances, and covariances are set to zero (Preacher & Selig, 2012). We can then simulate data using

$$\begin{pmatrix} a^* \\ B^* \end{pmatrix} \sim MVN\left[\begin{pmatrix} \hat{a} \\ \hat{B} \end{pmatrix}, \begin{pmatrix} \hat{\sigma}^2_{\hat{a}} & 0 \\ 0 & \hat{\sigma}^2_{\hat{B}} \end{pmatrix}\right]. \tag{16}$$

The sampling distribution can then be estimated by multiplying a^* and B^* and power is the proportion of asymmetric confidence intervals (e.g., 95%) that exclude no effect.

We again probed the functional relationships between each of the parameters and power under the Monte Carlo interval test with the same values (Figure 1). Overall, the relationships between power and each of the parameters were very similar to those of the previous tests.

Optimal sampling

The Monte Carlo interval test does directly draw on the error variance of the mediation effect but rather indirectly uses the error variances of the individual path coefficients to approximate the posterior distribution for the mediation effect. In addition, similar to the joint test, maximizing the power for each individual test does not necessarily maximize the power of the Monte Carlo interval test. As a result, closed-form expressions that identify the sampling plan that optimizes power under this test are not readily available. Despite this limitation, we can develop two simple approaches for approximating the optimal sampling scheme. In a first approach, we can use brute force to estimate the power under a grid of alternative sampling schemes. This approach is computationally expensive but straightforward.

In a second approach, we can leverage the similarities between the Monte Carlo interval test and the Sobel and joint tests to approximate the optimal sampling scheme. More specifically, although the sampling distribution of the mediation effect is often poorly approximated under

small sample sizes, the Sobel-based asymptotic approximation of the error variance of the mediation effect converges to the true error variance quickly. Similarly, prior research has consistently shown that the power curves for the joint and Monte Carlo interval tests are similar (e.g., Kelcey, Dong, Spybrook, & Cox, 2017). In these ways, the optimal sample allocation under the Sobel and/or joint tests potentially serve as good approximations of the optimal sample allocation for the Monte Carlo interval test.

We probed the efficacy of the Sobel- and joint-based approximations approach by comparing the relationships between the parameters and the optimal sampling plans identified by the brute force approach with those of the Sobel and joint tests. The results are plotted on top of those of the Sobel and joint tests in Figure 2. Overall the results suggested that the functional relationship between the parameters and the optimal number of students per school under the Monte Carlo interval test is a mixture of the forms presented by the Sobel and joint tests. For the outcome and mediator variance (Figure 2d and 2f), all three tests had very similar forms and implied optimal sample sizes. For the cost ratio and the a path coefficient, the Monte Carlo interval test most closely aligned with the joint test in terms of its functional form, though there was still minor variation in their optimal sampling plans. However, the relationship between the optimal sampling plan under the Monte Carlo interval test and the B path coefficient was fairly muted compared to the relationships for the joint and Sobel tests; the optimal sampling plan under the Monte Carlo interval test appeared to be a compromise between that of the Sobel and the joint plans (Figure 2b).

Lower-level mediation effects

When effects are constant across students and schools, we can conceptualize the overall mediation effect as a combination of components that flow through the student or school level. The second type of mediation effect we thus consider is the unique individual or lower-level mediation effect. A lower-level mediation effect probes the degree to which the effects of a school-level treatment on a student-level outcome are transmitted through the student-level component of the mediator (e.g., Pituch & Stapleton, 2012). When mediation is constant across students, the lower-level mediation effect (LME) can be described as

$$\text{LME} = ab_1. \tag{17}$$

For instance, the lower-level mediation effect describes the growth in student achievement that accumulates as a result of changes in an individual student's behavior generated by participation in the RHC program when holding constant schoolmates' behavior.

Sobel test

The Sobel test statistic for the lower-level mediation effect is

$$z_{ab_1}^{Sobel_{211}} = ab_1 / \sqrt{\sigma_{ab_1}^2} \tag{18}$$

where $\sigma_{ab_1}^2$ is the lower-level mediation effect error variance such that

$$\sigma_{ab_1}^2 = a^2 \left(\frac{\sigma_{Y|}^2}{(n_2 n_1 - n_2)\sigma_{M|}^2} \right) + b_1^2 \left(\frac{\tau_{M|}^2 + \sigma_{M|}^2 / n_1}{P(1-P)n_2} \right). \tag{19}$$

Like the overall mediation effect, we can further unpack the error variance as a function of common summary statistics. The resulting power to detect a lower-level mediation effect is then

$$P\left(z_{ab_1}^{Sobel} > z_{critical}\right) = 1-\Phi\left(z_{critical}-z_{ab_1}^{Sobel}\right) + \Phi\left(-z_{critical}-z_{ab_1}^{Sobel}\right). \quad (20)$$

Optimal sampling

Similar to the overall mediation effect, we can derive the sampling allocation that specifically optimizes the power to detect the lower-level mediation effect. The first-order derivative of the error variance of the lower-level mediation effect in terms of n_1 is

$$\frac{d\sigma_{ab_1}^2}{dn_1} = \frac{1}{c}(\Phi-\Gamma+\mathrm{H}). \quad (21)$$

with

$$\Phi = \frac{a^2(c_1+c_2)\left(b_1^2\left(R_{M_z^{L1}}^2-1\right)\sigma_M^2-R_{Y_z^{L1}}^2\sigma_Y^2+\sigma_Y^2\right)}{\left(R_{M_z^{L1}}^2-1\right)\sigma_M^2(n_1-1)^2}.$$

$$\Gamma = \frac{4b_1^2c_1\left(n_1\left(0.25a^2+\left(R_{M_z^{L2}}^2-1\right)\tau_M^2\right)+\left(R_{M_z^{L2}}^2-1\right)\sigma_M^2\right)}{n_1}.$$

$$\mathrm{H} = \frac{4\mathrm{b}_1^2\left(\mathrm{R}_{\mathrm{M}_z^{L2}}^2-1\right)\sigma_{\mathrm{M}}^2(\mathrm{c}_1\mathrm{n}_1+\mathrm{c}_2)}{\mathrm{n}_1^2}$$

Again, there is no closed-form solution but we can easily obtain the optimal number of students by solving this derivative numerically. This routine is implemented in the *R* package in PowerUpR.

Joint test

We can adapt the joint test to lower-level mediation effect by substituting the test for the overall association between the mediator and outcome (B) with that of its student-level complement (b_1). The tests for lower-level mediation effect are then

$$z_a = a/\sqrt{\sigma_a^2} \quad \text{and} \quad z_{b_1} = b_1/\sqrt{\sigma_{b_1}^2}. \quad (22)$$

The error variance for the a path is unchanged from the overall mediation effect while the error variance for the b_1 path can be estimated using

$$\sigma_{b_1}^2 = \frac{\sigma_{Y|}^2}{(n_2n_1-n_2)\sigma_{M|}^2}. \quad (23)$$

The power of the joint test to detect lower-level mediation is then determined with

$$\begin{aligned} P\left(|z_a| > z_{critical}\&|z_{b_1}| > z_{critical}\right) &= \left(1-z(z_{critical}-z_a)+z(-z_{critical}-z_a)\right) \\ &*\left(1-z(z_{critical}-z_{b_1})+z(-z_{critical}-z_{b_1})\right) \end{aligned} \quad (24)$$

Optimal sampling

As with the Sobel test, we can further identify the sampling plan under the joint test that maximizes the power to specifically detect the lower-level mediation effect. The derivative of the power function for the lower-level joint test yields

$$\frac{\partial P(|z_a| > z_{critical} \;\&\; |z_{b_1}| > z_{critical})}{\partial n_1} = \frac{1}{c}\exp(-z^2_{critical})\{(\frac{a(c_2\sigma^2_M(4-4R^2_{M^{L1}_z}) + c_1 n_1^2(a^2 + (4R^2_{M^{L2}_z}-4)\tau^2_M))}{10n_1^2(\frac{a}{D})^{1.5}}\exp(D^2)$$
$$(\exp(\frac{1}{2}(z-D)^2)-\exp(\frac{1}{2}(D+z)^2))(erfc(\frac{E-z}{\sqrt{2}})-erfc(\frac{E+z}{\sqrt{2}})-2)-\frac{\frac{1}{10}b_1(c_1+c_2)(b_1^2(R^2_{M^{L1}_z}-1)\sigma^2_M - R^2_{Y^{L1}_z}\sigma^2_Y + \sigma^2_Y)}{(R^2_{M^{L1}_z}-1)\sigma^2_M(n_1-1)^2(\frac{b_1}{E})^{1.5}}$$
$$\exp(E^2)(-erfc(\frac{\sqrt{2}}{2}D+\frac{z}{\sqrt{2}}) + erfc(\frac{\sqrt{2}}{2}D-\frac{z}{\sqrt{2}})-2)(\exp(\frac{1}{2}(z-E)^2)-\exp(\frac{1}{2}(E+z)^2))\}. \tag{25}$$

where *erfc* is the complementary error function and

$$D = \frac{a}{\left(-\frac{(c_1n_1+c_2)\left(n_1\left(a^2+4\left(R^2_{M^{L2}_z}-1\right)\tau^2_M\right)+4\left(R^2_{M^{L1}_z}-1\right)\sigma^2_M\right)}{cn_1}\right)^{0.5}} \text{ and } E = \frac{b_1}{\left(-\frac{(c_1n_1+c_2)\left(b_1^2\left(R^2_{M^{L1}_z}-1\right)\sigma^2_M - R^2_{Y^{L1}_z}\sigma^2_Y+\sigma^2_Y\right)}{c\left(R^2_{M^{L1}_z}-1\right)\sigma^2_M(n_1-1)}\right)^{0.5}} \tag{26ab}$$

The derivative is again somewhat verbose and without a simple closed-form solution. However, we can use this result to quickly identify the optimal number of students by solving the derivative numerically as implemented in the *R* package PowerUpR.

Monte Carlo interval test

The Monte Carlo interval test of lower-level mediation can also be extended in an analogous manner. We can draw samples from the posterior distribution using

$$\begin{pmatrix} a^* \\ b_1^* \end{pmatrix} \sim MVN\left[\begin{pmatrix} \hat{a} \\ \hat{b}_1 \end{pmatrix}, \begin{pmatrix} \hat{\sigma}^2_{\hat{a}} & 0 \\ 0 & \hat{\sigma}^2_{\hat{b}_1} \end{pmatrix}\right]. \tag{27}$$

The sampling distribution can then be estimated by multiplying a^* and b_1^* with power as the proportion of asymmetric confidence intervals (e.g., 95%) that exclude no effect.

Optimal sampling

As with the overall mediation effect, we can use the optimal sampling plan for the lower-level mediation effect under the Sobel test to approximate the optimal sampling scheme for the Monte Carlo interval test. We further outline the utility of this approach in our illustration section below.

Upper-level mediation effects

The final type of mediation effect we exam is the upper-level mediation effect (UME). The upper-level mediation effect outlines the contextual or environmental effect and examines the association of the school-means with achievement beyond the association linking individual student behavior and achievement (as captured by b_1). For instance, the contextual or upper-level mediation effect describes the improvement in student achievement that accrues as a result of changes in schoolmates' behaviors when holding constant individual student behavior.

Our current school-mean-centered models do not directly parameterize the upper-level mediation effect because the school-mean parameterization produces a student-level coefficient (b_1) that delineates the student-level-mediator-outcome association and a school-level coefficient (B) that lumps together the student-level and the school-level mediator-outcome association. The unique school-level associations can be obtained by changing the outcome model (2) such

that the student-level mediator values reflect the absolute standing (e.g., grand-mean centered) or the original mediator values of students within schools rather than the school-mean-centered values (Raudenbush & Bryk, 2002; Pituch & Stapleton, 2012). The uncentered outcome model is

$$\begin{aligned} Y_{ij} &= \beta_{0j} + b_1 M_{ij} + \beta_1\left(X_{ij} - \bar{X}_j\right) + \varepsilon_{ij}^{Y} \quad \varepsilon_{ij}^{Y} \sim N\left(0, \sigma_{Y|}^2\right) \\ \beta_{0j} &= \gamma_{00} + b_2 \bar{M}_j + c' T_j + \gamma_{01} W_j + \gamma_{02} \bar{X}_j + u_{0j}^{Y} \quad u_{0j}^{Y} \sim N\left(0, \tau_{Y|}^2\right) \end{aligned} . \tag{28}$$

In this parameterization, the school-level coefficient attached to the school-level mean mediator (b_2) now represents the specific contextual or unique school-level conditional association between the mediator and outcome.

Alternatively, we can retain the school-mean-centered parameterization originally presented (Equation 2) and obtain an estimate of the contextual or upper-level mediation effect by taking the difference of the overall and lower-level mediation effects. For the upper-level mediation effect (UME) this yields:

$$\text{UME} = ab_2 = a(B - b_1). \tag{29}$$

where b_2 could be estimated as the school-level coefficient using the uncentered model (28) or from the school-mean-centered model in Expression (3).

Sobel test

To track statistical power, we can draw on similar tests for the upper-level mediation effect. The Sobel test statistic can be extended to

$$z_{ab_2} = ab_2 / \sqrt{\sigma_{ab_2}^2} \tag{30}$$

where $\sigma_{ab_2}^2$ is the error variance of the upper-level mediation. This error variance can be traced by considering the relationships of the b_2 path with the B and b_1 paths. Specifically, the error variance of the upper-level mediation effect can be estimated as

$$\sigma_{ab_2}^2 = a^2\left(\sigma_B^2 + \sigma_{b_1}^2\right) + (B - b_1)^2\sigma_a^2 = a^2\left[\frac{\left(\tau_{Y|}^2 + \sigma_{Y|}^2/n_1\right)}{n_2\left(\tau_{M|}^2 + \sigma_{M|}^2/n_1\right)} + \frac{\sigma_{Y|}^2}{(n_2 n_1 - n_2)\sigma_{M|}^2}\right] + (B - b_1)^2 \frac{\tau_{M|}^2 + \sigma_{M|}^2/n_1}{P(1-P)n_2}. \tag{31}$$

Like the previous effects, we can reduce this expression by substituting in functions of key summary statistics for the conditional variances. Power under the Sobel test for upper-level mediation can then be estimated as

$$P\left(z_{ab_2}^{Sobel} > z_{critical}\right) = 1 - \Phi\left(z_{critical} - z_{ab_2}^{Sobel}\right) + \Phi\left(-z_{critical} - z_{ab_2}^{Sobel}\right). \tag{32}$$

Optimal sampling

Like previous developments, we can identify the sampling allocation that specifically optimizes the power to detect the upper-level mediation effect. The first-order derivative of the error variance of the upper-level mediation effect in terms of n_1 yields:

$$\frac{d\sigma_{ab_2}^2}{dn_1} = \frac{1}{c}\left(a^2(-\mathrm{T} + \Upsilon - \mathrm{K} + \Delta - \theta + \omega)\right) \tag{33}$$

with

$$\mathrm{T} = \frac{4(R^2_{M^{L1}_z}-1)\sigma^2_M(c_1n_1+c_2)(-2aBc'n_1+4B^2((R^2_{M^{L1}_z}-1)\sigma^2_M+(R^2_{M^{L2}_z}-1)\tau^2_Mn_1)+4b_1^2(R^2_{M^{L1}_z}-1)\sigma^2_M-c'^2n_1-4R^2_{Y^{L1}_z}s^2_Y-4R^2_{Y^{L2}_z}\tau^2_Yn_1+4\sigma^2_Y+4\tau^2_Yn_1)}{n_1(n_1(a^2+4(R^2_{M^{L2}_z}-1)\tau^2_M)+4(R^2_{M^{L1}_z}-1)\sigma^2_M)^2}$$

$$\Upsilon = \frac{4(c_1n_1+c_2)(B^2(R^2_{M^{L1}_z}-1)\sigma^2_M+b_1^2(R^2_{M^{L1}_z}-1)\sigma^2_M-R^2_{Y^{L1}_z}\sigma^2_Y+\sigma^2_Y)}{n_1(n_1(a^2+4(R^2_{M^{L2}_z}-1)\tau^2_M)+4(R^2_{M^{L1}_z}-1)\sigma^2_M)}$$

$$\mathrm{K} = \frac{c_1(-2aBc'n_1+4B^2((R^2_{M^{L1}_z}-1)\sigma^2_M+(R^2_{M^{L2}_z}-1)\tau^2_Mn_1)+4b_1^2(R^2_{M^{L1}_z}-1)\sigma^2_M-c'^2n_1-4R^2_{Y^{L1}_z}\sigma^2_Y-4R^2_{Y^{L2}_z}\tau^2_Yn_1+4\sigma^2_Y+4\tau^2_Yn_1)}{n_1(a^2+4(R^2_{M^{L2}_z}-1)\tau^2_M)+4(R^2_{M^{L1}_z}-1)\sigma^2_M}$$

$$\Delta = \frac{(c_1+c_2)(b_1^2(R^2_{M^{L1}_z}-1)\sigma^2_M-R^2_{Y^{L1}_z}\sigma^2_Y+\sigma^2_Y)}{(R^2_{M^{L1}_z}-1)\sigma^2_M(n_1-1)^2})$$

$$\theta = \frac{c_1(B-b_1)^2(n_1(a^2+(4R^2_{M^{L2}_z}-4)\tau^2_M)+(4R^2_{M^{L1}_z}-4)\sigma^2_M)}{n_1}$$

$$\omega = \frac{4(R^2_{M^{L1}_z}-1.)\sigma^2_M(B-b_1)^2(c_1n_1+c_2)}{n_1^2}$$

The optimal number of students is identified by finding a numerical solution that causes the derivative to be equal to zero. We implement this optimization in PowerUpR.

Joint test

Extending the joint test to the upper-level mediation effect, we can formulate the corresponding test statistics as

$$z_a = a/\sqrt{\sigma^2_a} \quad \text{and} \quad z_{b_2} = b_2/\sqrt{\sigma^2_{b_2}}. \tag{34ab}$$

The error variance for the b_2 path can be estimated using:

$$\sigma^2_{b_2} = \frac{\left(\tau^2_{Y|}+\sigma^2_{Y|}/n_1\right)}{n_2\left(\tau^2_{M|}+\sigma^2_{M|}/n_1\right)} + \frac{\sigma^2_{Y|}}{(n_2n_1-n_2)\sigma^2_{M|}}. \tag{35}$$

The power of the joint test to detect upper-level mediation is the calculated as

$$\begin{aligned} P\left(|z_a| > z_{critical} \& |z_{b_2}| > z_{critical}\right) &= \left(1-z\left(z_{critical}-z_a\right)+z\left(-z_{critical}-z_a\right)\right) \\ &*\left(1-z\left(z_{critical}-z_{b_2}\right)+z\left(-z_{critical}-z_{b_2}\right)\right). \end{aligned} \tag{36}$$

Optimal sampling

Once again, we can ascertain the sampling allocation that specifically optimizes the power to detect the upper-level mediation effect under the joint test. The first-order derivative of the error variance of the upper-level mediation effect in terms of n_1 is given in the appendix. As with the previous results, the optimal number of students is identified by finding a numerical solution that causes the derivative to be equal to zero. We implement this optimization in the *R* package PowerUpR.

Monte Carlo interval test

The Monte Carlo interval test for upper-level mediation can be extended in an analogous manner. We can simply use the mediator-outcome path and error variance that corresponds with contextual effect.

Illustration

Let us again consider the design of a multilevel mediation study that probes the degree to which the RHC program impacts student achievement by improving student behavior (mediator) as

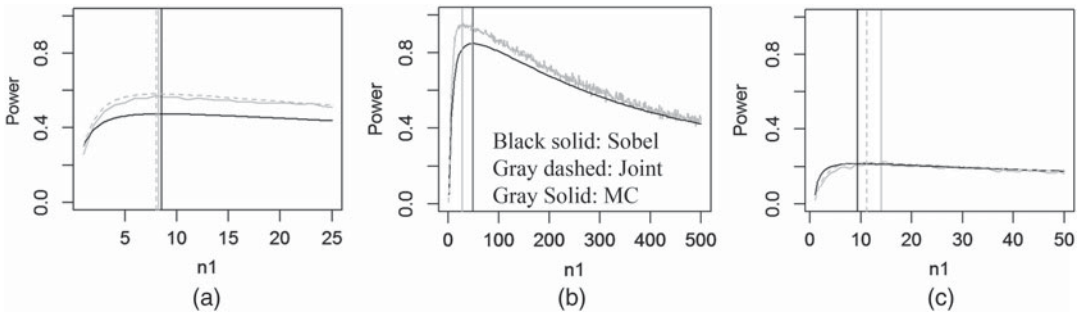

Figure 3. Power for the Sobel (black), joint (gray dashed), and Monte Carlo interval (gray line) tests for the (a) overall, (b) lower-level, and (c) upper-level mediation effect as a function of the number of students per school under specific cost constraints.

detailed through an overall mediation effect. We continue with the parameter values previously used. Specifically, we assign the unconditional school-level variances for the outcome and mediator to 0.20 (i.e., $\tau^2_M = \tau^2_Y = 0.20$) and the unconditional student-level variances to 0.80 (i.e., $\sigma^2_M = \sigma^2_Y = 0.80$); 10% of the variance is explained by covariates at each level for each response (i.e., $R^2_{Y^{L2}_z} = R^2_{M^{L2}_z} = R^2_{Y^{L1}_z} = R^2_{M^{L1}_z} = 0.10$ with z indicating the appropriate covariates for a response based on the models above); a treatment-mediator path coefficient of $a = 0.45$; an overall mediator-outcome association of $B = 0.35$ with 0.15 due to the student-level association ($b_1 = 0.15$); a direct effect of the treatment on the outcome of $c' = 0.05$; and half of the schools will receive the RHC program ($P = 0.50$). We can consider two complementary design questions that we can address with the results of our study: What is the sampling plan that yields the most powerful design in terms of detecting the overall mediation effect ($aB = 0.1575$) and under this budget what is the maximum amount of power under that sampling plan? Similarly, we ask, What are the most powerful sampling plans for the lower-level ($ab_1 = 0.0675$) and upper-level ($a(B\text{-}b_1) = ab_2 = 0.09$) mediation effects and how do these plans compare to that of the overall mediation effect?

The results of our analyses indicate that the optimal sampling plan for the overall mediation effect was to sample between $n_1 = 8$ and $n_1 = 9$ students per school with about $n_2 = 46$ schools yielding a power to detect the overall mediation effect of between 0.48 (Sobel) and 0.59 (joint/Monte Carlo; Figure 3). As a point of reference, the optimal sampling scheme for a main or total effect ($aB + c' = 0.2075$) was about $n_1 = 19$ students per school with 42 schools yielding a power level of 0.27. That is, in this particular example, under the respective optimal sampling plans, the power to detect the overall mediation exceeds the power for the main effect. If we adopted the optimal sampling plan for the overall mediation effect ($n_1 = 8$ and $n_1 = 9$), the power for the main effect would only decrease about 10%, to 0.25. If instead we adopted the sampling plan that was optimal for the main effect ($n_1 = 19$), the power for the overall mediation effect would drop by about 5%, to about 0.54, for the joint and Monte Carlo interval tests.

In terms of the lower-level mediation effect, the optimal sampling allocation according to the Sobel test was about $n_1 = 48$ students per school and $n_2 = 34$ schools; whereas, optimal for the Monte Carlo and joint tests is $n_1 = 28$ students per school with $n_2 = 39$ schools. When compared to the sample allocations that are optimal for the overall mediation effect ($n_1^{OME} = 8$–9), designs focusing on the lower-level mediation call for a larger number of students per school and yield much higher power (>0.8). Such a result is expected because holding constant the a path, the dominant sample size driving uncertainty for the overall mediation effect is the number schools; whereas, the dominant sample size driving uncertainty for the lower-level mediation effects is the number of students.

Now consider the degree to which improvements in achievement generated by participation in the RHC program is owing to, specifically, contextual improvements in the school (i.e., schoolwide

improvements in student behavior). The resulting optimal sampling allocation specific to detecting upper-level mediation ranges between $n_1 = 9$ students per school (Sobel) and about $n_1 = 14$ students per school (Monte Carlo) with the joint test splitting the differences. Similar to the optimal design for the overall mediation effect, the resulting school sample size is about $n_2 = 44$ to $n_2 = 46$ but the power for the upper-level mediation effect is much lower (0.2).

Discussion

Planning studies with the faculty to assess a more comprehensive set of effects—such as main and mediation effects—has become a core objective of education research. For example, the request for applications for Goal 3 Efficacy Studies from the Institute of Education Sciences strongly encourages applicants to include a clearly articulated theory of action, measures of the identified mediators, and mediation analyses (IES, 2017). A primary consideration in the effective and efficient planning of cluster-randomized studies is to identify the optimal allocation of resources across levels of the hierarchy and sample sizes that provide sufficiently powerful designs regarding targeted effects (e.g., Spybrook, Shi, & Kelcey, 2016).

In this study, we delineated three complementary types of mediation effects that trace the potential routes through which a treatment operates on an outcome. The mediation effects of interest in a given study depend on the target of the inference (Pituch & Stapleton, 2012). In many practical implementations, researchers will opt to plan studies using the overall mediation effect because it requires fewer assumptions, is statistically more defensible, and encompasses the mediation effects produced at either level when exposed to a treatment (i.e., the increments in the outcome produced by changes in either the student- or school-level mediators). For this reason, our analyses largely focused on the overall mediation effect and its behavior.

According to a broad assessment, our findings suggest that the power to detect the overall mediation effect was a complex function of multiple parameters. For some parameters, the nature of the relationships with power paralleled that of main effects whereas for other parameters the relationship broke down. For example, just as in the main effects analysis, the intraclass correlation coefficient of the outcome was negatively associated with power and adding covariates that accounted for that clustering improved power. In contrast, for instance, increases in the magnitude of the mediation effect (aB) were not necessarily associated with increases in power, as they are for increases in the main effect. Rather, the decomposition of the mediation effect (i.e., relative size of the a to B path) determines power and optimal sampling.

The results present several considerations to guide the design of multilevel mediation studies. Some of these guidelines parallel those of main effects—for example, adding covariates that explain outcome variance (regardless of their relation to the mediator) improves power; other results raise new design considerations. And even within these new considerations, some results suggest straightforward guidelines whereas others are much more complex. One simple consideration, for instance, is using the specific test of a mediation effect; prior literature would suggest identifying and using the most powerful test for a given setting.

A more complex consideration, however, arises when we examine, for instance, the implications of mediator variance. At first glance, a plausible strategy for improving power appears to be reducing the mediator variance explained by covariates at the school level because mediator variance has a negative relationship with power (Figure 1e). As noted earlier, further scrutiny suggests this is not a viable approach. Omitting covariates correlated with the mediator can improve power but doing so potentially undermines the viability of the sequential ignorability assumption. That is, because studies of mediation typically do not randomize mediator values, controlling for covariates in the outcome model is not optional but required to obtain an unbiased estimate of the mediation effect. Thus $R^2_{M^{L2}_z}$ is not truly a malleable parameter under

most designs. The result of omitting covariates correlated with the mediator to gain power would often result in a high level of power to detect a biased effect.

Another complex consideration is how we might balance the power and optimal sampling strategy for the main effect with that of a mediation effect. The results suggested that power for the overall mediation effect can be less than or greater than the power for main effects; it depends heavily on the specific parameter values. Our analysis also suggested that the optimal number of students per school for the overall mediation effect will usually be lower than that of the main effect and typically lower than about 25. In selecting between the resulting optimal sample allocations, researchers must establish study goals and priorities with these considerations in mind, potentially, privileging one effect, balancing both effects, or combining these considerations with other practical sampling constraints that arise in a given study. Future research may also look to derive formulas that identify a global optimal sampling plan—a sampling strategy that jointly optimizes the power to detect the main and mediation effects.

Similarly, our analyses constrained the behavior of a number of parameters that might be relaxed in future research. For example, our implementation assumed that the mediator and outcome variances and costs were similar across treatment groups (Shen & Kelcey, in review). What's more, our investigation assumed that researchers had a reasonably accurate a priori expectation as to parameter values (e.g., intraclass correlation coefficient). Future research may extend the current framework to relax such assumptions and probe the extent to which power and optimal sample allocation are sensitive to such assumptions (e.g., Manju, Candel, & Berger, 2014; Cox & Kelcey, in review).

In conclusion, our study provides expressions to trace the optimal sample allocation for and the power of multilevel mediation designs that can be easily implemented even before data collection. The resulting power and optimal sampling plan can be quickly approximated using the anticipated magnitudes of the path coefficients and common summary statistics. These results are intended to aid in effective and efficient design while identifying new considerations that are likely to arise in the design of multilevel mediation studies.

Note

1. Causal interpretation of indirect effects requires additional assumptions such as no downstream confounding variables. See for example, VanderWeele (2010) for details.

Funding

This project was partially supported by awards from the National Science Foundation [1552535, 1760884]. The opinions expressed herein are those of the authors and not the funding agency.

References

Aarons, G. A., Hurlburt, M., & Horwitz, S. M. (2011). Advancing a conceptual model of evidence-based practice implementation in public sectors. *Administration and Policy in Mental Health*, *38*(1), 4–23.

Beasley, T. (2014). Test of mediation: Paradoxical decline in statistical power as a function of mediator collinearity. *Journal of Experimental Education*, *82*(3), 283–306.

Cox, K., & Kelcey, B. (in press). Optimal sample allocation in group-randomized studies of multilevel mediation with a group-level mediator. *Journal of Experimental Education*.

Curenton, S. M., Dong, N., & Shen, X. (2015). Does aggregate school achievement mediate long-term outcomes for early childhood education participants? *Developmental Psychology*, *51*(7), 921–934.

Fleming, C. B., Harachi, T. W., Catalano, R. F., Haggerty, K. P., & Abbott, R. D. (2001). Assessing the effects of a school-based intervention on unscheduled school transfers during elementary school. *Evaluation Review*, *25*(6), 655–679.

Hayes, A. F., & Scharkow, M. (2013). The relative trustworthiness of inferential tests of the indirect effect in statistical mediation analysis: Does method really matter? *Psychological Science*, *24*(10), 1918–1927.
Hong, G., & Raudenbush, S. W. (2006). Evaluating kindergarten retention policy: A case study of causal inference for multilevel observational data. *Journal of the American Statistical Association*, *101*(475), 901–910.
Institute of Education Sciences [IES], U.S. Department of Education & National Science Foundation. (2013). *Common guidelines for education research and development* (NSF 13-126). Retrieved from http://ies.ed.gov/pdf/CommonGuidelines.pdf
Institute of Education Sciences [IES]. (2017). U.S. department of education institute of education sciences request for applications: Education research grants CFDA number: 84.305A.
Kelcey, B., Dong, N., Spybrook, J., & Cox, K. (2017). Statistical power for causally-defined indirect effects in group-randomized trials with individual-level mediators. *Journal of Educational and Behavioral Statistics*, *42*(5), 499–530.
Kelcey, B., Dong, N., & Spybrook, J. (2018). Sample size planning in cluster-randomized studies of multilevel mediation. *Prevention Science*,
Kelcey, B., & Phelps, G. (2013). Considerations for designing group randomized trials of professional development with teacher knowledge outcomes. *Educational Evaluation and Policy Analysis*, *35*(3), 370–390.
Kelcey, B., & Phelps, G. (2013). Strategies for improving power in school randomized studies of professional development. *Evaluation Review*, *37*(6), 520–554.
Kelcey, B., Hill, H., & Chin, M. (in press). Teacher mathematical knowledge, instructional quality, and student outcomes: A multilevel mediation analysis. *School Effectiveness & School Improvement*, *47*(1), 133–180.
Kelcey, B., Phelps, G., Spybrook, J., Jones, N., & Zhang, J. (2017). Designing large-scale multisite and cluster-randomized studies of professional development. *Journal of Experimental Education*, *85*(3), 389–410.
Kelcey, B., Dong, N., Spybrook, J., & Shen, Z. (2017). Experimental power for indirect effects in group-randomized studies with group-level mediators. *Multivariate Behavioral Research*, *52*(6), 699–719.
Kenny, D. A., & Judd, C. M. (2014). Power anomalies in testing mediation. *Psychological Science*, *25*, 334–339.
Kersting, N., Givvin, K., Thompson, B., Santagata, R., & Stigler, J. (2012). Measuring usable knowledge teachers' analyses of mathematics classroom videos predict teaching quality and student learning. *American Educational Research Journal*, *49*(3), 568–589.
Konstantopoulos, S. (2009). Incorporating cost in power analysis for three-level cluster-randomized designs. *Evaluation Review*, *33*(4), 335–357.
Kozlowski, S. W., & Klein, K. J. (2000). A multilevel approach to theory and research in organizations: Contextual, temporal, and emergent processes. In K. J. Klein & S. W. J. Kozlowski (Eds.), *Multilevel theory, research, and methods in organizations: Foundations, extensions, and new directions* (pp. 3–90). San Francisco: Jossey-Bass.
Kreft, I. G., de Leeuw, J., & Aiken, L. S. (1995). The effect of different forms of centering in hierarchical linear models. *Multivariate Behavioral Research*, *30*(1), 1–21.
Krull, J. L., & MacKinnon, D. P. (1999). Multilevel mediation modeling in group-based intervention studies. *Evaluation Review*, *23*(4), 418–444.
Krull, J. L., & MacKinnon, D. P. (2001). Multilevel modeling of individual and cluster level mediated effects. *Multivariate Behavioral Research*, *36*(2), 249–277.
Kunst, J., Fischer, R., Sidanius, J., & Thomsen, L. (2017). Preferences for group dominance track and mediate the effects of macro-level social inequality and violence across societies. *PNAS*, *114*(21), 5407–5412. Retrieved from https://doi. org/10.1073/pnas.1616572114
MacKinnon, D. P., Lockwood, C. M., & Williams, J. (2004). Confidence limits for the indirect effect: Distribution of the product and resampling methods. *Multivariate Behavioral Research*, *39*(1), 99–128.
Manju, M., Candel, M., & Berger, M. (2014). Sample size calculation in cost-effectiveness cluster r randomized trials: Optimal and maximin approaches. *Statistics in Medicine*, *33*(15), 2538–2553.
Nagengast, B., & Marsh, H. W. (2012). Big fish in little ponds aspire more: Mediation and cross-cultural generalizability of school-average ability effects on self-concept and career aspirations in science. *Journal of Educational Psychology*, *104*(4), 1033–1053.
Pituch, K., & Stapleton, L. (2012). Distinguishing between cross- and cluster-level mediation processes in the cluster randomized trial. *Sociological Methods & Research*, *41*(4), 630–670.
Preacher, K. J., & Selig, J. P. (2012). Advantages of Monte Carlo confidence intervals for indirect effects. *Communication Methods and Measures*, *6*(2), 77–98.
Raudenbush, S. W. (1997). Statistical analysis and optimal design for cluster randomized trials. *Psychological Methods*, *2*(2), 173–185.
Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods*. Thousand Oaks, CA: Sage.
Saarento, S., Boulton, A., & Salmivalli, C. (2015). Reducing bullying and victimization: Student- and classroom-level mechanisms of change. *Journal of Abnormal Child Psychology*, *43*(1), 61–76.

Shen, Z., & Kelcey, B. (in review). Optimal sample allocation under unequal costs in cluster-randomized trials. *Journal of Educational and Behavioral Statistics*, *28*(3), 231–248

Spybrook, J., Shi, R., & Kelcey, B. (2016). Progress in the past decade: An examination of the precision of cluster randomized trials funded by the U.S. institute of education sciences. *International Journal of Research & Method in Education*, *39*(3), 255–267.

Sobel, M. E. (1982). Asymptotic confidence intervals for indirect effects in structural equation models. *Sociological Methodology*, *13*, 290–312.

Stäbler, F., Dumont, H., Becker, M., & Baumert, J. (2017). What happens to the fish's achievement in a little pond? A simultaneous analysis of class-average achievement effects on achievement and academic self-concept. *Journal of Educational Psychology*, *109*(2), 191–207.

Talloen, W., Moerkerke, B., Loeys, T., De Naeghel, J., Van Keer, H., & Vansteelandt, S. (2016). Estimation of indirect effects in the presence of unmeasured confounding for the mediator-outcome relationship in a multilevel 2-1-1 mediation model. *Journal of Educational and Behavioral Statistics*, *41*(4), 359–391.

VanderWeele, T. J. (2010). Direct and indirect effects for neighborhood-based clustered and longitudinal data. *Sociological Methods & Research*, *38*(4), 515–544.

VanderWeele, T. J. (2015). *Explanation in causal inference: Methods for mediation and interaction*. New York, NY: Oxford University Press.

VanderWeele, T. J., & Vansteelandt, S. (2009). Conceptual issues concerning mediation, interventions and composition. *Statistics and Its Interface*, *2*(4), 457–468.

Wagner, L., & Ruch, W. (2015). Good character at school: Positive classroom behavior mediates the link between character strengths and school achievement. *Frontiers in Psychology*, *6*, 610–627.

Zhang, Z., Zyphur, M., & Preacher, K. (2009). Testing multilevel mediation using hierarchical linear models: Problems and solutions. *Organizational Research Methods*, *12*(4), 695–719.

Appendix

First-order derivative of power for the overall mediation effect under the joint test:

$$\frac{\partial P(|z_a| > z_{critical} \;\&\; |z_B| > z_{critical})}{\partial n_1} = \frac{1}{c}\exp^{-z^2_{critical}}\Big(\frac{1}{n_1^2\Big(-\frac{(c_2+c_1n_1)(4(R^2_{M^{L1}_z}-1)\sigma^2_M+(a^2+4(R^2_{M^{L2}_z}-1)\tau^2_M)n_1)}{cn_1}\Big)^{1.5}}\Big(\frac{a}{10}\Big)$$

$$\exp^{\frac{a^2cn_1}{(c_2+c_1n_1)(4(R^2_{M^{L1}_z}-1)\sigma^2_M+(a^2+4(R^2_{M^{L2}_z}-1)\tau^2_M)n_1)}}\big(\exp^{\frac{1}{2}(z_{critical}-G)^2}-\exp^{\frac{1}{2}(G+z_{critical})^2}\big)$$

$$(c_1(a^2+(4R^2_{M^{L2}_z}-4)\tau^2_M)n_1^2+c_2(4-4R^2_{M^{L1}_z})\sigma^2_M)\,(erfc(\frac{H-z_{critical}}{\sqrt{2}})-erfc(\frac{H+z_{critical}}{\sqrt{2}})-2)-$$

$$\{\frac{2B}{5}\exp(-\frac{B^2c(4(R^2_{M^{L1}_z}-1)\sigma^2_M+(a^2+4(R^2_{M^{L2}_z}-1)\tau^2_M)n_1)}{(c_2+c_1n_1)(-4((R^2_{M^{L1}_z}-1)\sigma^2_M+(R^2_{M^{L2}_z}-1)\tau^2_Mn_1)B^2+2ac'n_1B-4b_1^2(R^2_{M^{L1}_z}-1)\sigma^2_M+4R^2_{Y^{L1}_z}\sigma^2_Y-4\sigma^2_Y+c'^2n_1+4R^2_{Y^{L2}_z}\tau^2_Yn_1-4\tau^2_Yn_1)})$$

$$[\exp^{\frac{1}{2}(z_{critical}-H)^2}-\exp^{\frac{1}{2}(H+z_{critical})^2}][0.5Bc_1c'n_1^2a^3+((c_1(1-R^2_{M^{L2}_z})\tau^2_Mn_1^2+c_2(R^2_{M^{L1}_z}-1)\sigma^2_M)B^2+0.25c_1c'^2n_1^2-c_1\tau^2_Yn_1^2+c_1R^2_{Y^{L2}_z}\tau^2_Yn_1^2+$$

$$b_1^2c_2(R^2_{M^{L1}_z}-1)\sigma^2_M+c_2\sigma^2_Y-c_2R^2_{Y^{L1}_z}\sigma^2_Y)a^2+Bc'(c_2(2R^2_{M^{L1}_z}-2)\sigma^2_M+c_1n_1((4R^2_{M^{L1}_z}-4)\sigma^2_M+(2R^2_{M^{L2}_z}-2)\tau^2_Mn_1))a+$$

$$c_2((4R^2_{M^{L2}_z}R^2_{M^{L1}}-4R^2_{M^{L1}}-4R^2_{M^{L2}_z}+4)\sigma^2_M\tau^2_Mb_1^2+c'^2(R^2_{M^{L1}}-1)\sigma^2_M+4R^2_{M^{L2}_z}\sigma^2_Y\tau^2_M-4R^2_{M^{L2}_z}R^2_{Y^{L1}_z}\sigma^2_Y\tau^2_M+4R^2_{Y^{L1}_z}\sigma^2_Y\tau^2_M-$$

$$4\sigma^2_Y\tau^2_M-4R^2_{M^{L1}}\sigma^2_M\tau^2_Y+4R^2_{M^{L1}}R^2_{Y^{L2}_z}\sigma^2_M\tau^2_Y-4R^2_{Y^{L2}_z}\sigma^2_M\tau^2_Y+4\sigma^2_M\tau^2_Y)+c_1((-4(1-R^2_{M^{L1}_z})^2\sigma^4_M+(-8R^2_{M^{L2}_z}R^2_{M^{L1}}+8R^2_{M^{L1}}+8R^2_{M^{L2}_z}-8)\tau^2_Mn_1\sigma^2_M-$$

$$4(1-R^2_{M^{L2}_z})^2\tau^4_Mn_1^2)B^2-4b_1^2(1-R^2_{M^{L1}_z})^2\sigma^4_M{}^2-c'^2\tau^2_Mn_1^2+c'^2R^2_{M^{L2}_z}\tau^2_Mn_1^2-4R^2_{M^{L2}_z}\tau^2_M\tau^2_Yn_1^2+4R^2_{M^{L2}_z}R^2_{Y^{L2}_z}\tau^2_M\tau^2_Yn_1^2-4R^2_{Y^{L2}_z}\tau^2_M\tau^2_Yn_1^2+4\tau^2_M\tau^2_Yn_1^2-$$

$$4R^2_{M^{L1}_z}\sigma^2_M\sigma^2_Y+4R^2_{M^{L1}_z}R^2_{Y^{L1}_z}\sigma^2_M\sigma^2_Y-4R^2_{Y^{L1}_z}\sigma^2_M\sigma^2_Y+4\sigma^2_M\sigma^2_y-2c'^2\sigma^2_m\sigma^2_m+2c'^2R^2_{M^{L1}_z}\sigma^2_mn_1-8R^2_{M^{L1}_z}\sigma^2_m\tau^2_yn_1+$$

$$8.R^2_{M^{L1}_z}R^2_{Y^{L2}_z}\sigma^2_M\tau^2_Yn_1-8R^2_{Y^{L2}_z}\sigma^2_M\tau^2_Yn_1+8\sigma^2_M\tau^2_Yn_1)]\;[-erfc(\frac{\sqrt{2}}{2}G+\frac{z_{critical}}{\sqrt{2}})+erfc(\frac{\sqrt{2}}{2}G-\frac{z_{critical}}{\sqrt{2}})-2]\}/$$

$$\{((4R^2_{M^{L1}_z}-4)\sigma^2_M+(a^2+(4R^2_{M^{L2}_z}-4)\tau^2_M)n_1)^2$$

$$(\frac{(c_2+c_1n_1)(-4((R^2_{M^{L1}_z}-1)\sigma^2_M+(R^2_{M^{L2}_z}-1)\tau^2_Mn_1)B^2+2ac'n_1B-4b_1^2(R^2_{M^{L1}_z}-1)\sigma^2_M+4R^2_{Y^{L1}_z}\sigma^2_Y-4\sigma^2_Y+c'^2n_1+4R^2_{Y^{L2}_z}\tau^2_Yn_1-4\tau^2_Yn_1)}{c(4(R^2_{M^{L1}_z}-1)\sigma^2_M+(a^2+4(R^2_{M^{L2}_z}-1)\tau^2_M)n_1)})^{1.5}\})$$

(A1)

where *erfc* as the complementary error function and

$$H=\frac{B}{\left(\frac{(c_2+c_1n_1)\left(-4\left(\left(R^2_{M^{L1}_z}-1\right)\sigma^2_M+\left(R^2_{M^{L2}_z}-1\right)\tau^2_Mn_1\right)B^2+2ac'n_1B-4b_1^2\left(R^2_{M^{L1}_z}-1\right)\sigma^2_M+4R^2_{Y^{L1}_z}\sigma^2_Y-4\sigma^2_Y+c'^2n_1+4R^2_{Y^{L2}_z}\tau^2_Yn_1-4\tau^2_Yn_1\right)}{c\left(4\left(R^2_{M^{L1}_z}-1\right)\sigma^2_M+\left(a^2+4\left(R^2_{M^{L2}_z}-1\right)\tau^2_M\right)n_1\right)}\right)^{0.5}} \tag{A2}$$

and

$$G = \frac{a}{\left(-\frac{(c_2+c_1 n_1)\left(4\left(R^2_{M_z^{L1}}-1\right)\sigma_M^2+\left(a^2+4\left(R^2_{M_z^{L2}}-1\right)\tau_M^2\right)n_1\right)}{cn_1}\right)^{0.5}} \tag{A3}$$

First-order derivative of power for the upper-level mediation effect under the joint test:

$$\frac{\partial P(|z_a| > z_{critical} \;\&\; |z_{b_2}| > z_{critical})}{\partial n_1} =$$

$$\frac{1}{2c}\Bigg(\frac{1}{n_1^2\left(-\frac{(c_2+c_1 x)(4(R^2_{M_z^{L1}}-1)\sigma_M^2+(a^2+4(R^2_{M_z^{L2}}-1)\tau_M^2)n_1)}{cn_1}\right)^{1.5}}\frac{a}{5}\exp\left(\frac{a^2 cn_1}{(c_2+c_1 n_1)(4(R^2_{M_z^{L1}}-1)\sigma_M^2+(a^2+4(R^2_{M_z^{L2}}-1)\tau_M^2 m)n_1)}-z^2_{critical}\right)$$

$$\left(\exp\left(\frac{1}{2}\left(z_{critical}-\frac{a}{\left(-\frac{(c_2+c_1 n_1)(4(R^2_{M_z^{L1}}-1)\sigma_M^2+(a^2+4(R^2_{M_z^{L2}}-1)\tau_M^2)n_1)}{cn_1}\right)^{0.5}}\right)^2\right)-\exp\left(\frac{1}{2}\left(\frac{a}{\left(-\frac{(c_2+c_1 n_1)(4(R^2_{M_z^{L1}}-1)\sigma_M^2+(a^2+4(R^2_{M_z^{L2}}-1)\tau_M^2)n_1)}{cn_1}\right)^{0.5}}+z\right)^2\right)\right)$$

$$(c_1(a^2+(4R^2_{M_z^{L2}}-4)\tau_M^2)n_1^2+c_2(4-4R^2_{M_z^{L1}})\sigma_M^2)$$

$$\left(erfc\left(\frac{\frac{B-b_1}{(c_2+c_1 n_1)\left(\frac{-4\left(R^2_{M_z^{L1}}-1\right)\sigma_M^2+\left(R^2_{M_z^{L2}}-1\right)\tau_M^2 n_1)B^2+2ac' n_1 B-4b_1^2\left(R^2_{M_z^{L1}}-1\right)\sigma_M^2+4R^2_{Y_z^{L1}}\sigma_Y^2-4\sigma_Y^2+c'^2 n_1+4R^2_{Y_z^{L2}}\tau_Y^2 n_1-4\tau_Y^2 n_1}{4\left(R^2_{M_z^{L1}}-1\right)\sigma_M^2+\left(a^2+4\left(R^2_{M_z^{L2}}-1\right)\tau_M^2\right)n_1}-\frac{\left(R^2_{M_z^{L1}}-1\right)\sigma_M^2 b_1^2-R^2_{Y_z^{L1}}\sigma_Y^2+\sigma_Y^2}{\left(R^2_{M_z^{L1}}-1\right)\sigma_M^2(n_1-1)}\right)}-z_{critical}}{\left(\frac{(c_2+c_1 n_1)\left(\frac{-4\left(R^2_{M_z^{L1}}-1\right)\sigma_M^2+\left(R^2_{M_z^{L2}}-1\right)\tau_M^2 n_1)B^2+2ac' n_1 B-4b_1^2\left(R^2_{M_z^{L1}}-1\right)\sigma_M^2+4R^2_{Y_z^{L1}}\sigma_Y^2-4\sigma_Y^2+c'^2 n_1+4R^2_{Y_z^{L2}}\tau_Y^2 n_1-4\tau_Y^2 n_1}{4\left(R^2_{M_z^{L1}}-1\right)\sigma_M^2+\left(a^2+4\left(R^2_{M_z^{L2}}-1\right)\tau_M^2\right)n_1}-\frac{\left(R^2_{M_z^{L1}}-1\right)\sigma_M^2 b_1^2-R^2_{Y_z^{L1}}\sigma_Y^2+\sigma_Y^2}{\left(R^2_{M_z^{L1}}-1\right)\sigma_M^2(n_1-1)}\right)}{c}\right)^{0.5}}{\sqrt{2}}\right)-\right.$$

$$erfc\left(\frac{\frac{B-b_1}{(c_2+c_1 n_1)\left(\frac{-4\left(R^2_{M_z^{L1}}-1\right)\sigma_M^2+\left(R^2_{M_z^{L2}}-1\right)\tau_M^2 n_1)B^2+2ac' n_1 B-4b_1^2\left(R^2_{M_z^{L1}}-1\right)\sigma_M^2+4R^2_{Y_z^{L1}}\sigma_Y^2-4\sigma_Y^2+c'^2 n_1+4R^2_{Y_z^{L2}}\tau_Y^2 n_1-4\tau_Y^2 n_1}{4\left(R^2_{M_z^{L1}}-1\right)\sigma_M^2+\left(a^2+4\left(R^2_{M_z^{L2}}-1\right)\tau_M^2\right)n_1}-\frac{\left(R^2_{M_z^{L1}}-1\right)\sigma_M^2 b_1^2-R^2_{Y_z^{L1}}\sigma_Y^2+\sigma_Y^2}{\left(R^2_{M_z^{L1}}-1\right)\sigma_M^2(n_1-1)}\right)}+z_{critical}}{\left(\frac{(c_2+c_1 n_1)\left(\frac{-4\left(R^2_{M_z^{L1}}-1\right)\sigma_M^2+\left(R^2_{M_z^{L2}}-1\right)\tau_M^2 n_1)B^2+2ac' n_1 B-4b_1^2\left(R^2_{M_z^{L1}}-1\right)\sigma_M^2+4R^2_{Y_z^{L1}}\sigma_Y^2-4\sigma_Y^2+c'^2 n_1+4R^2_{Y_z^{L2}}\tau_Y^2 n_1-4\tau_Y^2 n_1}{4\left(R^2_{M_z^{L1}}-1\right)\sigma_M^2+\left(a^2+4\left(R^2_{M_z^{L2}}-1\right)\tau_M^2\right)n_1}-\frac{\left(R^2_{M_z^{L1}}-1\right)\sigma_M^2 b_1^2-R^2_{Y_z^{L1}}\sigma_Y^2+\sigma_Y^2}{\left(R^2_{M_z^{L1}}-1\right)\sigma_M^2(n_1-1)}\right)}{c}\right)^{0.5}}{\sqrt{2}}\right)-2)+$$

$$((B-b1)\Bigg(\frac{1}{5}\exp\left(-\frac{1}{2}\left(\frac{\frac{B-b_1}{(c_2+c_1 n_1)\left(\frac{-4\left(R^2_{M_z^{L1}}-1\right)\sigma_M^2+\left(R^2_{M_z^{L2}}-1\right)\tau_M^2 n_1)B^2+2ac' n_1 B-4b_1^2\left(R^2_{M_z^{L1}}-1\right)\sigma_M^2+4R^2_{Y_z^{L1}}\sigma_Y^2-4\sigma_Y^2+c'^2 n_1+4R^2_{Y_z^{L2}}\tau_Y^2 n_1-4\tau_Y^2 n_1}{4\left(R^2_{M_z^{L1}}-1\right)\sigma_M^2+\left(a^2+4\left(R^2_{M_z^{L2}}-1\right)\tau_M^2\right)n_1}-\frac{\left(R^2_{M_z^{L1}}-1\right)\sigma_M^2 b_1^2-R^2_{Y_z^{L1}}\sigma_Y^2+\sigma_Y^2}{\left(R^2_{M_z^{L1}}-1\right)\sigma_M^2(n_1-1)}\right)}+z_{critical}}{\left(\frac{(c_2+c_1 n_1)\left(\frac{-4\left(R^2_{M_z^{L1}}-1\right)\sigma_M^2+\left(R^2_{M_z^{L2}}-1\right)\tau_M^2 n_1)B^2+2ac' n_1 B-4b_1^2\left(R^2_{M_z^{L1}}-1\right)\sigma_M^2+4R^2_{Y_z^{L1}}\sigma_Y^2-4\sigma_Y^2+c'^2 n_1+4R^2_{Y_z^{L2}}\tau_Y^2 n_1-4\tau_Y^2 n_1}{4\left(R^2_{M_z^{L1}}-1\right)\sigma_M^2+\left(a^2+4\left(R^2_{M_z^{L2}}-1\right)\tau_M^2\right)n_1}-\frac{\left(R^2_{M_z^{L1}}-1\right)\sigma_M^2 b_1^2-R^2_{Y_z^{L1}}\sigma_Y^2+\sigma_Y^2}{\left(R^2_{M_z^{L1}}-1\right)\sigma_M^2(n_1-1)}\right)}{c}\right)^{0.5}}\right)^2\right)-$$

$$\frac{1}{5}\exp\left(-\frac{1}{2}\left(\frac{-(B-b_1)}{(c_2+c_1 n_1)\left(\frac{-4\left(R^2_{M_z^{L1}}-1\right)\sigma_M^2+\left(R^2_{M_z^{L2}}-1\right)\tau_M^2 n_1)B^2+2ac' n_1 B-4b_1^2\left(R^2_{M_z^{L1}}-1\right)\sigma_M^2+4R^2_{Y_z^{L1}}\sigma_Y^2-4\sigma_Y^2+c'^2 n_1+4R^2_{Y_z^{L2}}\tau_Y^2 n_1-4\tau_Y^2 n_1}{4\left(R^2_{M_z^{L1}}-1\right)\sigma_M^2+\left(a^2+4\left(R^2_{M_z^{L2}}-1\right)\tau_M^2\right)n_1}-\frac{\left(R^2_{M_z^{L1}}-1\right)\sigma_M^2 b_1^2-R^2_{Y_z^{L1}}\sigma_Y^2+\sigma_Y^2}{\left(R^2_{M_z^{L1}}-1\right)\sigma_M^2(n_1-1)}\right)}+z_{critical}}{\left(\frac{(c_2+c_1 n_1)\left(\frac{-4\left(R^2_{M_z^{L1}}-1\right)\sigma_M^2+\left(R^2_{M_z^{L2}}-1\right)\tau_M^2 n_1)B^2+2ac' n_1 B-4b_1^2\left(R^2_{M_z^{L1}}-1\right)\sigma_M^2+4R^2_{Y_z^{L1}}\sigma_Y^2-4\sigma_Y^2+c'^2 n_1+4R^2_{Y_z^{L2}}\tau_Y^2 n_1-4\tau_Y^2 n_1}{4\left(R^2_{M_z^{L1}}-1\right)\sigma_M^2+\left(a^2+4\left(R^2_{M_z^{L2}}-1\right)\tau_M^2\right)n_1}-\frac{\left(R^2_{M_z^{L1}}-1\right)\sigma_M^2 b_1^2-R^2_{Y_z^{L1}}\sigma_Y^2+\sigma_Y^2}{\left(R^2_{M_z^{L1}}-1\right)\sigma_M^2(n_1-1)}\right)}{c}\right)^{0.5}}\right)^2\right)\right)$$

$$\left(\frac{(c_1+c_2)((R^2_{M_z^{L1}}-1)\sigma_M^2 b_1^2-R^2_{Y_z^{L1}}\sigma_Y^2+\sigma_Y^2)}{(R^2_{M_z^{L1}}-1)\sigma_M^2(n_1-1)^2}-\frac{c_1(4((R^2_{M_z^{L1}}-1)\sigma_M^2+(R^2_{M_z^{L2}}-1)\tau_M^2 n_1)B^2-2ac' n_1 B+4b_1^2(R^2_{M_z^{L1}}-1)\sigma_M^2-4R^2_{Y_z^{L1}}\sigma_Y^2+4\sigma_Y^2-c'^2 n_1-4R^2_{Y_z^{L2}}\tau_Y^2 n_1+4\tau_Y^2 n_1)}{4(R^2_{M_z^{L1}}-1)\sigma_M^2+(a^2+4(R^2_{M_z^{L2}}-1)\tau_M^2 m)n_1}+\right.$$

$$\frac{4((R^2_{M_z^{L1}}-1)\sigma_M^2 B^2+b_1^2(R^2_{M_z^{L1}}-1)\sigma_M^2-R^2_{Y_z^{L1}}\sigma_Y^2+\sigma_Y^2)(c_2+c_1 n_1)}{n_1(4(R^2_{M_z^{L1}}-1)\sigma_M^2+(a^2+4(R^2_{M_z^{L2}}-1)\tau_M^2)n_1)}-$$

$$\left.\frac{4((R^2_{M_z^{L1}}-1)\sigma_M^2(c_2+c_1 n_1)(4((R^2_{M_z^{L1}}-1)\sigma_M^2+(R^2_{M_z^{L2}}-1)\tau_M^2 n_1)B^2-2ac' n_1 B+4b_1^2(R^2_{M_z^{L1}}-1)\sigma_M^2-4R^2_{Y_z^{L1}}\sigma_Y^2+4\sigma_Y^2-c'^2 n_1-4R^2_{Y_z^{L2}}\tau_Y^2 n_1+4\tau_Y^2 n_1)}{n_1(4((R^2_{M_z^{L1}}-1)\sigma_M^2+(a^2+4(R^2_{M_z^{L2}}-1)\tau_M^2)n_1)^2}\right)$$

$$\left(erfc\left(\frac{\frac{\sqrt{2}}{2}a}{\left(-\frac{(c_2+c_1 n_1)(4(R^2_{M_z^{L1}}-1)\sigma_M^2+(a^2+4(R^2_{M_z^{L2}}-1)\tau_M^2)n_1)}{cn_1}\right)^{0.5}}+\frac{z}{\sqrt{2}}\right)-erfc\left(\frac{\frac{\sqrt{2}}{2}a}{\left(-\frac{(c_2+c_1 n_1)(4(R^2_{M_z^{L1}}-1)\sigma_M^2+(a^2+4(R^2_{M_z^{L2}}-1)\tau_M^2)n_1)}{cn_1}\right)^{0.5}}-\frac{z}{\sqrt{2}}\right)+2)\right)$$

$$/\left(\frac{(c_2+c_1 n_1)\left(\frac{-4\left(R^2_{M_z^{L1}}-1\right)\sigma_M^2+\left(R^2_{M_z^{L2}}-1\right)\tau_M^2 n_1)B^2+2ac' n_1 B-4b_1^2\left(R^2_{M_z^{L1}}-1\right)\sigma_M^2+4R^2_{Y_z^{L1}}\sigma_Y^2-4\sigma_Y^2+c'^2 n_1+4R^2_{Y_z^{L2}}\tau_Y^2 n_1-4\tau_Y^2 n_1}{4\left(R^2_{M_z^{L1}}-1\right)\sigma_M^2+\left(a^2+4\left(R^2_{M_z^{L2}}-1\right)\tau_M^2\right)n_1}-\frac{\left(R^2_{M_z^{L1}}-1\right)\sigma_M^2 b_1^2-R^2_{Y_z^{L1}}\sigma_Y^2+\sigma_Y^2}{\left(R^2_{M_z^{L1}}-1\right)\sigma_M^2(n_1-1)}\right)}{c}\right)^{1.5}\Bigg) \tag{A4}$$