CrossMark

Sample Size Planning for Cluster-Randomized Interventions Probing Multilevel Mediation

Ben Kelcey[1] · Jessaca Spybrook[2] · Nianbo Dong[3]

Abstract

Multilevel mediation analyses play an essential role in helping researchers develop, probe, and refine theories of action underlying interventions and document how interventions impact outcomes. However, little is known about how to plan studies with sufficient power to detect such multilevel mediation effects. In this study, we describe how to prospectively estimate power and identify sufficient sample sizes for experiments intended to detect multilevel mediation effects. We outline a simple approach to estimate the power to detect mediation effects with individual- or cluster-level mediators using summary statistics easily obtained from empirical literature and the anticipated magnitude of the mediation effect. We draw on a running example to illustrate several different types of mediation and provide an accessible introduction to the design of multilevel mediation studies. The power formulas are implemented in the *R* package *PowerUpR* and the *PowerUp* software (causalevaluation.org).

Keywords Mediation · Power · Indirect effects · Multilevel models · Sample size determination

A common line of inquiry in prevention science studies is to investigate the intermediate behaviors or attitudes that transmit the effects of an intervention on an outcome (Gottfredson et al. 2015). In many instances, these types of investigations take place within multilevel contexts and, as a result, include the examination of not only individual behaviors but also the collective behavior of individuals within a cluster (e.g., Krull and MacKinnon 2001). The predominant approach in these types of investigations is to test the sequence of relationships connecting an intervention, mediator, and outcome through a multilevel mediation framework (Pituch and Stapleton 2012; Zhang et al. 2009). The goals of such studies are typically to unpack the processes whereby an intervention impacts an individual- or cluster-level mediator variable lying along the intervention-outcome pathway and how those changes in the mediator are then translated into changes in an outcome.

Such inquiries advance scientific theory by building a multilayered body of evidence regarding if, how, and why an intervention impacts an outcome. Investigation of the transmission channels of an intervention can help develop component-specific evidence regarding the theory of action guiding that intervention as well as evidence regarding the effectiveness of the entire system or intervention (e.g., Gottfredson et al. 2015). For these reasons, (multilevel) mediational analyses have been widely used across many disciplines (e.g., Williams and Glisson 2014).

Similar to other types of inquiry, the quality and utility of inferences drawn from a mediation study are heavily contingent upon the maturity of the intervention theory and the corresponding study design. For instance, mediation studies are typically limited by the extent to which theory has identified and delineated alternative causal connections that would compete with or supplant the working theory of action (e.g., sequential ignorability). Well-designed multilevel experiments that measure the mediator and outcome in a temporal sequence along with potential confounding variables, however, can to some extent relax these limitations and improve the quality of inferences regarding intervention and mediation effects (e.g., Pituch and Stapleton 2012). For instance, there is a growing body of literature that describes how addressing the core assumptions that support mediation inferences (e.g., sequential ignorability, stable unit treatment value assumption) and employing specific design features (e.g., random assignment, observation of potential confounding variables, sufficient sample sizes) can be leveraged to understand and improve interpretation, reduce bias, increase precision, and buttress the overall quality of evidence produced through

✉ Ben Kelcey
   ben.kelcey@gmail.com

[1] University of Cincinnati, Cincinnati, OH 45221, USA

[2] Western Michigan University, Kalamazoo, MI 49008, USA

[3] University of North Carolina, Chapel Hill, NC 27599, USA

mediation studies (e.g., VanderWeele 2010; Kelcey and Shen, in press; Cox and Kelcey, in press).

One key to ensuring that these types of investigations will be well-positioned to produce clear evidence regarding hypotheses is identifying sample sizes that will provide a sufficient level of precision to detect anticipated effects (e.g., Spybrook et al. 2016). Although both methodological research on multilevel mediation and its applications in substantive areas continue to be quite active, a significant gap in this literature is determining sample sizes that provide a sufficient level of power to detect mediation effects if they exist. From a prospective design view, there is little research or software on how to calculate the power to detect multilevel mediation effects or the degree to which it is feasible to design multilevel studies with sufficient power under sample sizes that are typical for a particular area.

In this study, we develop simple calculations and software to estimate power for multilevel mediation effects under cluster-randomized designs. We detail and illustrate the sensitivity with which two-level hierarchical designs can detect the production of the core mediation paths— the path linking the treatment and the mediator and the path linking the mediator and the outcome. The scope of our analyses includes statistical power when considering a cluster-level or individual-level mediator. Tracing the sensitivity of hierarchical designs to detect mediation effects is essential to the principal aims of mediation studies—to assess theories of action—because the sensitivity of these designs constrains the types of evidence tests of multilevel mediation can bring to bear on theories of action under particular sample sizes.

We detail these methods as follows. First, we set up an example to provide an applied context to our work and to serve as a common illustration across different types of mediators and mediation effects. Second, we use this example to outline the statistical power to detect multilevel mediation when the mediator is a cluster-level variable. Third, we extend this work and example to designs that probe multilevel mediation with an individual-level mediator. At the end of each section, we present illustrations of the power calculations. We end with a discussion.

Illustrative Example We begin by outlining a running example in order to describe mediation concepts and their application. We consider the design of a multilevel mediation study that assesses the potential pathways through which health practitioner participation in a training program on effective patient consultation (intervention) affects their use of an evidence-based counseling approach termed motivational interviewing (outcome). Our running example focuses on hierarchical designs that nest practitioners within community health organizations and assign

those organizations (and all their respective practitioners) to a training program or control condition. We then investigate the extent to which the impact of participation in an organization-assigned training program (cluster-level intervention) on the motivational interviewing practices of practitioners (individual-level outcome) operates through different types of intermediate variables (cluster- or individual-level mediators).

Motivational interviewing is a patient-centered counseling approach that encourages practitioners to proactively share and discuss with patients the potential problems and risks associated with a particular behavioral practice (Schwartz, 2010). Prior literature has demonstrated consistent evidence of the critical role this type of interviewing can play in improving patient outcomes. This example is also part of a wider and growing literature focused on how researchers can improve the dissemination and adoption of evidence-based behavioral practices by healthcare practitioners in, for example, community health organizations.

Theories of action for this example have typically focused on models of social diffusion and implementation to outline how practices are taken up, implemented, and sustained by organizations (Schwartz, 2010). However, there is a growing recognition of how little is known about the specific processes of translating empirical evidence about the efficacy of certain practices into widespread use by healthcare practitioners in social organizations.

Within this context, we consider two different types of processes and mediating variables. The first example targets an organizational process and examines the mediating role of a cluster- or organization-level variable that describes variation among organizational units and is constant among practitioners within organizations. In this literature, prevalent types of organizational mediators are those describing organizations' climate or the conditions under which practitioners operate within particular organizations (e.g., Williams & Glisson, 2014). We take up one possible operationalization of this construct—directors' readiness for change as reported by the director of each organization (TCU 2005). The theory of action typically suggests that training in the benefits of an effective practice cultivates directors' readiness to support and implement the changes, and this director-guided support subsequently promotes the use of the practice.

Because director readiness for change is measured using the reports of the directors at each organization, it represents an organization-level mediator in that it captures only differences among organizations and not differences among practitioners within organizations. As a result, with this first type of cluster- or organization-level mediator, we consider a form of 2-2-1 or upper-level mediation where the number acronym indicates the levels of the treatment (level 2), mediator (level 2), and outcome

(level 1). In 2-2-1 upper-level mediation, we assess the flow of treatment effects from an intervention assigned at the organization-level through an organization-level mediator to a practitioner-level outcome.

The second example process we consider involves an individual- or practitioner-level mediating variable that tracks variation among individual practitioners and thus is not constant within organizations. One common practitioner-level mediator in this literature is practitioners' attitudes toward the evidence-based practice. Theories of action often suggest that adoption of a new practice is preceded by changes in attitudes regarding that practice. With a practitioner-level mediator, the analysis is often referred to as 2-1-1 mediation because it considers how the impact of an organization-level treatment (level 2) impacts a practitioner-level (level 1) mediator in ways that yield improvements in a practitioner-level outcome (level 1).

Practitioner attitudes represent an individual-level mediator in that they capture differences among practitioners. However, practitioner attitudes can also be used to document differences among organizations in terms of context and these differences may represent additional mediator pathways. For instance, within many organizations, the collective sentiment of practitioners at that organization toward a practice may further cultivate or complicate the use of that practice. For this reason, many theories of action intentionally incorporate and leverage the role of cluster dynamics in shaping outcomes.

The most common operationalization of collective sentiment or attitude in this context is to consider its organization-level mean. Within this regard, mediation analyses consider how exposure to a training program affects the use of a practice by probing how exposure produces changes in individual attitudes and how exposure alters the organizational context brought about by changes in the collection of individual attitudes. As a result, it is common in studies with individual-level mediators to detail the flow of treatment effects as they pass from a treatment assigned at the organization-level through both practitioner-level attitudes and collective attitudes to an outcome measured at the practitioner-level. For this reason, our analyses with practitioner-level mediators consider overall, lower- and upper-level mediation.

2-2-1 Mediation

We first outline statistical power for experimental designs tracking organization-level mediators (i.e., 2-2-1 mediation; see Fig. 1a). Our analyses concentrate on designs that assign organizations at random to a control or treatment condition (T) and assess its impact on a continuous practitioner-level outcome through a continuous organization-level mediator (M). In our example, this may translate into assigning organizations

to a training program or a control condition with the intention of assessing its impact on eventual use of that practice as it operates through changes in director readiness for change. Our multilevel mediation formulation is

$$M_j = \pi_0 + aT_j + \pi_1 W_j + \pi_2 \overline{X}_j + \varepsilon_j^M \qquad \varepsilon_j^M \sim N\left(0, \sigma_{M|}^2\right) \quad (1a)$$

$$Y_{ij} = \beta_{0j} + \beta_1\left(X_{ij} - \overline{X}_j\right) + \beta_2 V_{ij} + \varepsilon_{ij}^Y \qquad \varepsilon_{ij}^Y \sim N\left(0, \sigma_{Y|}^2\right) \quad (1b)$$

$$\beta_{0j} = \gamma_{00} + bM_j + c'T_j + \gamma_{01}W_j + \gamma_{02}\overline{X}_j + u_{0j} \qquad u_{0j} \sim N\left(0, \tau_{Y|}^2\right) \quad (1c)$$

For the mediation model (1a), we use M_j as the mediator for organization j, W_j as an organization-level covariate with π_1 as its coefficient, X_{ij} as a practitioner-level covariate for practitioner i in organization j that potentially varies across individuals with organizations but also across organizations, $\overline{X}_j$ as the aggregate of a practitioner-level covariate with π_2 as its coefficient, T_j as the randomly allocated treatment with a capturing the treatment's impact on the mediator, and ε_j^M as the error term with conditional normal distribution $\varepsilon_j^M \sim N\left(0, \sigma_{M|}^2\right)$.

For our healthcare example, the mediation model targets how exposure to a training program changes a director's readiness for change. Our formulation additionally incorporates the possibility of organization-level covariates and aggregated practitioner covariates. Organization-level covariates might include, for instance, directors' pre-treatment levels of readiness for change whereas aggregated practitioner covariates might include the average pretreatment use of the targeted practice by practitioners at an organization. Random assignment of the treatment ensures that, in expectation, there are no variables that confound the association between treatment assignment and mediator values. However, including covariates that explain variation in the mediator can often improve the statistical precision with which we can estimate the treatment-mediator relationship (a path) and ultimately the mediation effect.

For the outcome model (1bc), we use a linear mixed model with Y_{ij} as the outcome for practitioner i in organization j, X_{ij} as a practitioner-level covariate with coefficient β_1, V_{ij} as a practitioner-level covariate that only varies among individuals within organizations with coefficient β_2, and ε_{ij}^Y as the level 1 error term with conditional normal distribution $\varepsilon_{ij}^Y \sim N\left(0, \sigma_{Y|}^2\right)$. At the organization-level, γ_{01} and γ_{02} are the path coefficients for the organization-level covariate W and aggregate of the practitioner-level covariate X, b as the conditional relationship between the mediator and the outcome, c' as the direct effect of the treatment,

Fig. 1 Path diagrams for **a** 2-2-1
mediation, **b** 2-1-1 mediation
under no centering and
decomposition approach, **c** 2-1-1
mediation under cluster-mean
centering and decomposition
approach, and **d** 2-1-1 mediation
under the cluster-level mediation
only approach
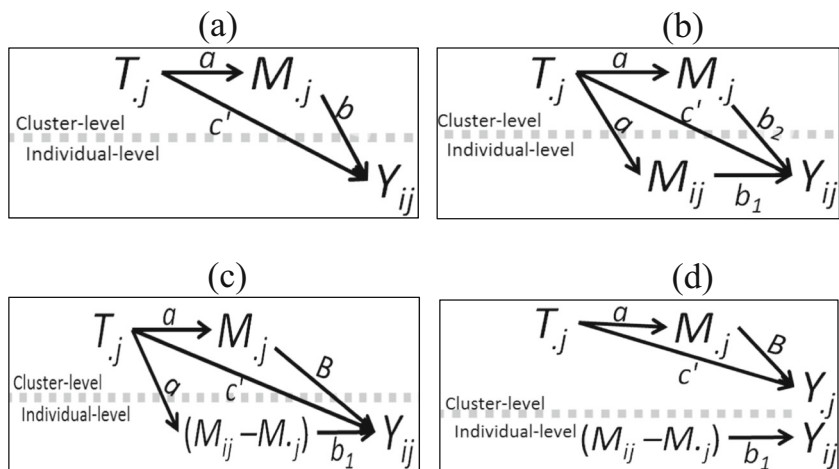
and u_{0j} as the organization-level random intercept with
conditional normal distribution $u_{0j} \sim N\left(0, \tau_{Y|}^2\right)$.

Using our running example, the outcome model delin-
eates how changes in director readiness for change are as-
sociated with improved use of motivational interviewing by
practitioners, while controlling for exposure to the training
program. Like the mediation model, the outcome model also
incorporates the possibility of conditioning on covariates so
that estimates control for confounding variables. However,
in the outcome model, inclusion of covariates will typically
be required in order to address the potential for confound-
ing—the possibility that the observed mediator–outcome
association is attributable to another variable.

When assumptions regarding causal inference and the
model (see VanderWeele 2010; Kelcey et al. 2017 for
details) are met, the 2-2-1 multilevel mediation effect
(ME) is typically estimated using the product of the
intervention-mediator (a) and mediator–outcome (b) paths
$ME_{221} = ab$.[1] The mediation effect maps out the impact of
the intervention on the outcome as it acts through changes
in the organization-level mediator. From a practical per-
spective, this mediation effect captures how changes in
director readiness for change brought about by participa-
tion in a training program manifest as changes in practi-
tioner use of motivational interviewing.

We note that our presentation makes a simplifying as-
sumption in that does not incorporate a treatment-mediator
interaction. The implication of this assumption is that ex-
posure to the treatment is assumed to impact the outcome
through changes in the mediator but not through changes in
the magnitude of the mediator's relationship with the out-
come. This assumption can be relaxed by including inter-
actions and the accompanying extensions regarding

indirect effects, their error variance and the resulting power
are detailed in the literature (Kelcey et al. 2017).

Effect Size If we standardize the outcome and mediator to
have an unconditional mean of zero and unit variance (i.e.,
$\sigma_M^2 = 1, \tau_Y^2 + \sigma_Y^2 = \rho_Y + (1-\rho_Y) = 1$), the organization-
level outcome variance can now be interpreted as the
intracluster correlation coefficient (ρ_Y). The magnitudes of
the a (treatment-mediator) path and c' (direct effect of the
treatment on the outcome) path are positioned on a stan-
dardized mean difference scale whereas the b (mediator–
outcome) path is located on a standardized regression coef-
ficient for an organization-level variable. Alternatively, on
the basis of prior multilevel literature, (Stapleton, Pituch,
and Dion, 2015) proposed three different mediation effect
sizes that better suit clustered experiments: (a) a single-
cluster effect size, (b) a multi-cluster or broad population
effect size, and (c) a cluster-level effect size.

The single-cluster effect size (d_w) focuses on changes
in individual outcomes within a cluster and can be used to
describe the standardized mediation effect if the treatment
were implemented within a single site. To obtain the
single-cluster effect size, we can divide the mediation ef-
fect by the unconditional individual-level outcome stan-
dard deviation ($d_w = ab/\sigma_Y$).

In contrast, the multi-cluster or broad population effect
size (d_T) concentrates on changes in individuals when the
treatment is implemented across many clusters in a popula-
tion. It is estimated by dividing the mediation effect by the
square root of the total outcome variance that remains after
conditioning on the treatment only ($\vartheta_{Y|T}$). This total variance
can be obtained as $\vartheta_{Y|T} = \tau_Y^2\left(1-\frac{P(1-P)}{\tau_Y^2}(ab+c')^2\right) +\sigma_Y^2$.
Last, the cluster-level effect size describes effects as they
impact cluster-level outcomes (e.g., average use of interviewing
across all practitioners at an organization). It is estimated
by dividing the mediation effect by the square root of

[1] Core assumptions include (a) stable unit treatment value assumption, (b)
sequential ignorability, (c) consistency, (d) no downstream confounders, and
(e) no treatment-by-mediator interaction.

the cluster-level outcome variance that remains after conditioning on the treatment only ($\tau_{Y \mid T}$). The variance can be estimated as $\tau_{Y|T} = \tau_Y^2 \left(1 - \frac{P(1-P)}{\tau_Y^2}(ab + c')^2 \right)$.

Power Analysis Our approach focuses on tracking statistical power under the Monte Carlo interval test. The relative and absolute accuracy of this test as well as its type 1 error rate and power have been widely tested and shown to perform very well across multiple criteria (e.g., Preacher and Selig 2012; Kelcey et al. 2017; Kelcey et al. 2017). Critical to the planning phase (as opposed to analytic phase), this test can be employed before data collection.

The Monte Carlo interval test is a resampling-based alternative to Sobel-like tests. The primary advantage of the Monte Carlo interval test is that it uses resampling to track the sampling distribution of the mediation effect under finite sample sizes rather than asymptotic approximations of that distribution (Preacher and Selig 2012). To obtain samples from the distribution of the mediation effect, the Monte Carlo interval test employs the primary path coefficient estimates and their error variances. With a sufficient number of draws, we can estimate the sampling distribution of the estimated mediation effect and test a hypothesis of no mediation effect by assessing whether specific confidence intervals include zero. 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 consistently robust performance relative to bootstrap-based and other methods. However, the Monte Carlo interval test additionally provides two key advantages. First, it can be employed without access to full data or during the design phase when no data are available. Second, estimates of the power of this test do not require complex and computationally intensive resampling or Monte Carlo simulation because the test and confidence intervals can be reduced to resampling only simple sufficient statistics.

Under maximum likelihood estimation of the parameters, we apply the Monte Carlo confidence interval test by sampling from a multivariate normal distribution for the estimated path coefficients. The vector of means is set to the anticipated values for the a and b path coefficients and the covariance matrix is set to have diagonal terms equal to the corresponding path error variances and zero covariance (Kelcey et al. 2017). To track power in relation to hypothesized path coefficients in the planning stages, we derived the implied error variances as functions of three types of summary statistics that are common in the literature: (a) the primary path coefficients (i.e., a, b, c'), (b) the variance explained in the outcome at each level by covariates ($R^2_{Y^{L2}_{W,\overline{X}}}$ and $R^2_{Y^{L1}_{\overline{X},V}}$) and the variance in the mediator ($R^2_{M_{W,\overline{X}}}$) explained by covariates, and (c) the unconditional

intracluster correlation coefficient of the outcome (ρ_Y)—see Table 1 for a summary of the parameters as well as examples. Similar to the anticipated magnitude of a main or total intervention effect, values for the first type of summary statistic (primary path coefficients) will typically be developed by considering the nature of the intervention, the sensitivity of the mediator and outcome to intervention effects, and empirical benchmarks for effect sizes regarding the mediator and outcome (e.g., Phelps, Kelcey, Liu, and Jones 2016; Kelcey et al. 2017). Values for the second and third type (variance explained by covariates and variance decompositions) will commonly come from prior empirical literature (e.g. Kelcey et al. 2017; Kelcey and Phelps 2013).

We then approximate the sampling distribution of a mediation effect with the production of a^* and b^* with power as the proportion of asymmetric intervals (e.g., 95%) that exclude zero. The result is that by specifying the mediation path coefficients, the variance explained by covariates, and the variance decomposition of the outcome, we can quickly estimate the sample size needed to achieve a specific power level for detecting mediation effects.

Illustration

As previously outlined, consider the design of a multilevel mediation study that assesses the extent to which an organization-wide training program (intervention) impacts practitioner use of motivational interviewing (outcome) by cultivating directors' readiness for change (mediator). Assume that based on prior literature, we anticipate that the intracluster correlation coefficient for the outcome will be approximately 0.30 (ρ_Y) and that the covariates will explain 60% of the outcome and mediation variance at each level ($R^2_{Y^{L2}_{W,\overline{X}}} = R^2_{Y^{L1}_{\overline{X},V}} = R^2_{M_{W,\overline{X}}} = 0.60$). Such an assertion suggests that prior empirical research documenting the prognostic capacity of the covariates absent the intervention would point to the covariates reducing the unconditional mediator and outcome variances by 60% at each level. If we standardize the outcome and mediator, the resulting (initial) conditional mediator variance would become $\sigma_{M|}^2 = 0.40$ whereas the conditional outcome variance at the organization- and practitioner-level would decrease to $\tau_{Y|}^2 = 0.12$ and $\sigma_{Y|}^2 = 0.28$ (consideration of the mediator and treatment subsequently further reduces these conditional variances). Further hypothesize that the mediation path is composed of $a = 0.6$ and $b = 0.3$ and the direct effect is $c' = 0.10$. The conditional mediator variance is now $\sigma_{M|}^2 = 0.31$ whereas the conditional outcome variance at the organization- and practitioner-level would decrease to $\tau_{Y|}^2 = 0.0725$ and $\sigma_{Y|}^2 = 0.28$. Let us assume that we intend to assign 50% of the organizations to the intervention

Table 1 Summary of parameters needed to estimate power in a 2-2-1 study

Parameter	Parameter value	Label	Example
Alpha level (α)	0.05	Probability of a type 1 error	Rejecting the null of no mediation effect when an effect truly does not exist
Effect size: a	0.60	Treatment-mediator path coefficient as standardized mean difference	Impact of organizational training on directors' readiness for change
Effect size: b	0.30	Mediator–outcome path coefficient as standardized regression coefficient	Conditional association between directors' readiness for change and practitioner practices
Effect size: c'	0.10	Direct treatment-outcome path coefficient as standardized mean difference	Conditional association between organizational training and practitioner practices
Outcome intraclass correlation (ρ_Y)	0.30	Proportion of variance in outcome that is between clusters	Proportion of variation in practitioner practices attributable to differences among organizations
R^2_{YL1}	0.60	Proportion of outcome variance at individual-level explained by individual-level covariates	Proportion of variance among practitioners in their practice explained by pre-intervention use of similar practices (same as $R^2_{Y_{L1}|\bar{X},V}$)
R^2_{YL2}	0.60	Proportion of outcome variance at cluster-level explained by cluster-level covariates (excluding the treatment and mediator)	Proportion of variance among organizations in their practice explained by directors' readiness for change prior to the intervention (same as $R^2_{Y_{L2}|W,\bar{X}}$)
R^2_M	0.60	Proportion of mediator variance explained by cluster-level covariates (excluding the treatment)	Proportion of variance among organizations in their directors' readiness for change that is explained by directors' pre-intervention readiness for change (same as $R^2_{M_{W,\bar{X}}}$)
n (average cluster size)	30	Mean number of individuals per cluster (harmonic mean recommended)	Number of practitioners per organization sampled (same as n_1)
J (sample size [# of clusters])	42	Total number of clusters (total = treatment + control)	Number of intervention and control organizations sampled (same as n_2)
P	0.50	Proportion of level 2 units randomized to treatment: $J_T / (J_T + J_C)$	Proportion of organizations in the intervention condition
Mediation effect (ME)	0.18	ME = a*b	The effect of training on practitioner practices that operates through directors' readiness for change
Power (1-β) for ME: Sobel test	0.71	Statistical power to detect the mediation effect (ME) for two-tailed test using Sobel test	Power under an asymptotic approximation
Power (1-β) for ME: Joint test	0.89	Statistical power to detect the mediation effect (ME) for two-tailed test using joint test	Power under concurrent tests of the individual path coefficients
Power (1-β) for ME: MC	0.91	Statistical power to detect the mediation effect (ME) for two-tailed test using Monte Carlo simulation (5000 replications)	Power under the resampling-based Monte Carlo interval test

($P = 0.50$) while sampling 30 practitioners within each organization (n_1). How many organizations must we sample to have a 90% chance of detecting the mediation effect?

We conducted the power analyses using the *PowerUp-Mediator* Excel-based macro (Dong, Kelcey, Spybrook, and Maynard 2017). Figure 2 provides a screenshot of the user interface for the 2-2-1 module—it provides a label and a brief description of the parameters needed to evaluate power and then asks users to input the relevant values of these parameters. It provides three statistical tests of mediation. The first test is the historical Sobel test based on asymptotic approximation. The second test is the joint test that draws inferences based on the significance of the individual path coefficients—support for mediation is inferred when both paths are non-zero. Power for both of these tests can be estimated using

closed-form solutions and, as a result, once a parameter value has been modified, the power is automatically updated for the Sobel and joint tests. The third test implemented is the Monte Carlo interval test—because it is a resampling-based test, estimates of power are not updated upon modification of a parameter value. Rather, power estimates for the Monte Carlo interval test are updated on-demand by pushing the Run MC button located at the bottom of the screen.

Our application of the approach outlined above suggested that as few as 42 organizations (21 treatment, 21 control) would be sufficient to achieve a power level of about 0.90. That is, if we sampled 42 organizations and assigned 21 of them to participate in the intervention while observing the remaining 21 organizations without intervening, we would have almost a 90% chance of detecting the mediation effects if they exist.

Model CRA2-2-1: Power Calculator for Two-Level Cluster Random Assignment Design — Treatment at Level 2 and Mediator at Level 2		
Parameter Values		Comments
Alpha Level (α)	0.05	Probability of a Type I error
Effect Size: a	0.60	Treatment-mediator path coefficient as standardized mean difference
Effect Size: b	0.30	Mediator-outcome path coefficient as standardized regression coefficient
Effect Size: c'	0.10	Direct treatment-outcome path coefficient as standardized mean difference
Intraclass Correlation (ρ)	0.30	Proportion of variance in outcome that is between clusters
R^2_{YL1}	0.60	Proportion of outcome variance at individual-level explained by individual-level covariates
R^2_{YL2}	0.60	Proportion of outcome variance at cluster-level explained by cluster-level covariates (excluding the treatment and mediator)
R^2_M	0.60	Proportion of mediator variance explained by cluster-level covariates (excluding the treatment)
n (Average Cluster Size)	30	Mean number of individuals per cluster (harmonic mean recommended)
J (Sample Size [# of Clusters])	42	Total number of clusters (total=treatment +contol)
P	0.50	Proportion of level 2 units randomized to treatment: $J_T / (J_T + J_C)$
Mediaiton Effect (ME)	0.180	ME = a*b
Power (1-β) for ME: Sobel Test	0.728	Statistical power to detect the Mediation Effect (ME) for two-tailed test using Sobel test
Power (1-β) for ME: Joint Test	0.886	Statistical power to detect the Mediation Effect (ME) for two-tailed test using joint test
Power (1-β) for ME: MC	0.906	Statistical power to detect the Mediation Effect (ME) for two-tailed test using Monte Carlo simulation (2000 replications)

Run MC

Model CRA2-1-1: Power Calculator for Two-Level Cluster Random Assignment Design — Treatment at Level 2 and Mediator at Level 1		
Assumptions		Comments
Alpha Level (α)	0.05	Probability of a Type I error
Effect Size: a	0.60	Treatment-mediator path coefficient as standardized mean difference
Effect Size: b_l	0.10	Mediator-outcome path coefficient at the individual-level as a standardized regression coefficient
Effect Size: B	0.50	Overall mediator-outcome path coefficient as a standardized regression coefficient (overall coef=individual coef+cluster coef)
Effect Size: c'	0.10	Direct treatment-outcome path coefficient as standardized mean difference
Intraclass Correlation (ρ_M)	0.30	Proportion of variance in the mediator (M) that is between clusters
Intraclass Correlation (ρ_Y)	0.30	Proportion of variance in the outcome (Y) that is between clusters
$R^2_{M^{L1}}$	0.60	Proportion of mediator variance at individual-level explained by individual-level covariates
$R^2_{M^{L2}}$	0.60	Proportion of mediator variance at cluster-level explained by cluster-level covariates (excluding the treatment)
$R^2_{Y^{L1}}$	0.60	Proportion of outcome variance at individual-level explained by individual-level covariates (excluding mediator)
$R^2_{Y^{L2}}$	0.60	Proportion of outcome variance at cluster-level explained by cluster-level covariates (excluding the treatment and mediator)
n (Average Cluster Size)	30	Mean number of individuals per cluster (harmonic mean recommended)
J (Sample Size [# of Clusters])	80	Total number of clusters (total=treatment +contol)
P	0.50	Proportion of level 2 units randomized to treatment: $J_T / (J_T + J_C)$
Lower-Level Mediaiton Effect (LME)	0.06	LME = a * b1
Upper-Level Mediaiton Effect (UME)	0.24	UME = a * (B-b1)
Overall Mediaiton Effect (OME)	0.30	OME = a * B
Power (1-β) for LME: Sobel Test	0.995	Statistical power to detect the lower-level mediation effects for two-tailed test using Sobel test
Power (1-β) for LME: Joint Test	0.998	Statistical power to detect the lower-level mediation effects for two-tailed test using joint test
Power (1-β) for LME: MC	0.998	Statistical power to detect the lower-level mediation effects for two-tailed test using Monte Carlo simulation (2000 replications)
Power (1-β) for UME: Sobel Test	0.690	Statistical power to detect the upper-level mediation effects for two-tailed test using Sobel test
Power (1-β) for UME: Joint Test	0.693	Statistical power to detect the upper-level mediation effects for two-tailed test using joint test
Power (1-β) for UME: MC	0.697	Statistical power to detect the upper-level mediation effects for two-tailed test using Monte Carlo simulation (2000 replications)
Power (1-β) for OME: Sobel Test	0.866	Statistical power to detect the overall mediation effects for two-tailed test using Sobel test
Power (1-β) for OME: Joint Test	0.874	Statistical power to detect the overall mediation effects for two-tailed test using joint test
Power (1-β) for OME: MC	0.880	Statistical power to detect the overall mediation effects for two-tailed test using Monte Carlo simulation (2000 replications)

Run MC

Fig. 2 Screenshots of the *PowerUp-Mediator* software for the 2-2-1 module (top) and 2-1-1 module (bottom)

As a comparison, the sample size needed to achieve a 90% level to detect the "main" or total effect of $c = ab + c' = 0.28$ would be nearly double (72 organizations) the sample size needed for detecting mediation. The results of this illustration are not generalizable—the relationship between the power to detect the total effect (c) and the power to detect the mediation effect (ab) depend heavily on the decomposition of the mediation effect. In some configurations, the power to detect the total effect will exceed that of the mediation and in other configurations the converse will prevail.

2-1-1 Mediation

We next consider multilevel mediation analyses that examine the extent to which the impact of an organization-level treatment on a practitioner-level outcome is mediated by a practitioner-level mediator. Within our running example, we consider practitioner attitudes toward the evidence-based practice as the mediator. The resulting theory of action we would hypothetically intend to test would investigate the extent to which changes in practitioner attitudes (mediator) generated by participation in the training program (intervention) manifest as an increase in the use of motivational interviewing (outcome).

We begin with the typical multilevel formulation that employs a system of linear mixed path models (e.g., Pituch and Stapleton 2012). Our initial model draws on centering practitioner variables within organizations (or organization-mean centering) because this is the most common in the literature and is useful for disentangling mediation effects across levels. However, as we subsequently outline, using raw values or

centering on the grand mean across organizations will yield equal parameter estimates in the case of random intercept formulations. For a practitioner-level mediator, our mediator model is

$$M_{ij} = \pi_{0j} + \pi_1\left(X_{ij} - \overline{X}_j\right) + \pi_2 V_{ij} + \varepsilon_{ij}^M \qquad \varepsilon_{ij}^M \sim N\left(0, \sigma_{M|}^2\right) \qquad (3a)$$

$$\pi_{0j} = \zeta_{00} + aT_j + \zeta_{01}W_j + \zeta_{02}\overline{X}_j + u_{0j}^M \qquad u_{0j}^M \sim N\left(0, \tau_{M|}^2\right) \qquad (3b)$$

M_{ij} represents the mediator value for practitioner i in organization j, X_{ij} is a practitioner-level covariate (with π_1 as its path coefficient) that varies across individuals and organizations, V_{ij} is a practitioner-level covariate (with π_2 as its path coefficient) that varies only across individuals, $\overline{X}_j$ is the organization-level variable or mean aggregate of the practitioner-level variable (with ζ_{02} as its path coefficient), T_j is the treatment assignment with path coefficient a, ε_{ij}^M as the error term, and u_{0j}^M as the organization-specific random effects. Applied to our running example, the a path in this equation maps out how participation in a training program yields changes in practitioner attitudes toward the targeted practice.

The outcome model also parallels previous models such that

$$Y_{ij} = \beta_{0j} + b_1\left(M_{ij} - \overline{M}_j\right) + \beta_1\left(X_{ij} - \overline{X}_j\right) + \beta_2 V_{ij} + \varepsilon_{ij}^Y \qquad (4a)$$

$$\varepsilon_{ij}^Y \sim N\left(0, \sigma_{Y|}^2\right)$$

$$\beta_{0j} = \gamma_{00} + B\overline{M}_j + c' T_j + \gamma_{01}W_j + \gamma_{02}\overline{X}_j$$

$$+ u_{0j}^Y \qquad u_{0j}^Y \sim N\left(0, \tau_{Y|}^2\right) \qquad (4b)$$

We use the same notation as above and add $M_{ij} - \overline{M}_j$ as the organization-centered practitioner-level mediator with coefficient b_1, $\overline{M}_j$ as the mean of the mediator in organization 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 1 error terms. Returning to our substantive application, the B path coefficient delineates how changes in practitioner attitudes are (conditionally) correlated with their use of motivational interviewing (see below for further discussion).

Mediation Effect There are two predominant approaches to describing the flow of the intervention effects to the outcome through the mediator in the literature when interventions are assigned to clusters. In the first approach, mediation is allowed to operate through both the cluster- and individual-level variables when considering an individual-level mediator (e.g., Pituch and Stapleton 2012; VanderWeele 2010; Krull and MacKinnon 2001). This approach thus considers the extent to

which the intervention produces changes in the outcome by acting on both the individual-level mediator values (i.e., differences among practitioners within an organization) and in the cluster-level aggregate of the mediator (i.e., differences among organizations in context). As a result, it allows for multiple mediation pathways, each with a potentially different interpretation and different magnitude. Put differently, this approach asserts that an intervention can produce changes in the mediator that modify both the environment of a cluster and the individuals within a cluster in ways that influence the outcome. The result of this approach is that the (overall) mediation effect can be decomposed into a component attributable to the collective/contextual changes in the mediator for all individuals in a cluster and a component attributable to changes in the mediator values of specific individuals (see Fig. 1b, c).

In the second approach, analysts only consider mediation arising at the cluster-level (e.g., Zhang et al. 2009). The argument supporting this philosophy is that a cluster-assigned treatment can only be correlated with the cluster-level variation in the mediator. As a result, the treatment-mediator covariance can be subsequently correlated only with the cluster-level outcome variance (see Fig. 1d). Put differently, because individuals in the same cluster receive the same intervention, separating the extent to which the intervention works through the individual- versus cluster-level is not identified without making additional assumptions.

As a result, although these views appear to oppose each other, the practical difference reduces to whether one invokes additional assumptions to decompose the (overall) mediation effect into contextual and individual components or not. We further unpack these differences below but from a conceptual standpoint, the contextual- and individual-level mediation view considers the overall mediation effect and subsequently offers to descriptively decompose it into a unique cluster-level component (contextual) and a unique individual-level component. In contrast, the cluster-level mediation only view considers the exact same overall mediation effect but simply chooses not to decompose. As a result, the mathematics of estimation and statistical power are identical for the two approaches—one simply needs to decide if s/he is interested in only the overall mediation effect or interested in the overall mediation effect as well as its decomposition into a unique cluster-level component and a unique individual-level component. For this reason, our study accommodates both approaches and considers all three types of mediation effects researchers might examine when an individual-level mediator is of interest.

Overall Mediation Effects

To unpack these complementary types of mediation effects, let us consider three categories of mediation effects. The first

category we focus on is the overall mediation effect—the effect that is common across both of the aforementioned approaches and is not in dispute. The overall mediation effect is composed of the unique contextual (upper-level) and individual (cross-level) mediation effects (described below). That is, we use the overall mediation effect to describe how any changes in attitudes (be it individual and/or collective) brought about by an organization's participation in a training program manifest as changes in practitioner use of motivational interviewing. Under the centering within organization approach (as in expression (4ab)), the coefficient (B) attached to the organization mediator mean ($\overline{M}$) represents the sum of the mediator's practitioner-level relationship with the outcome (b_1) and the mediator's unique contextual relationship with the outcome (b_2; see below). That is, the B coefficient captures the total (individual plus contextual) influence of the mediator on the outcome as it operates through the practitioner-level mediator values and the mean scalar function of the organization mediator values. As a result, the product of the a and B coefficients represents the overall mediation effect of the treatment on the outcome as it operates through the individual mediator and the mean of the organization mediator values. This overall mediation effect will typically be the effect of primary interest in designing and analyzing a study regardless of the philosophical approach.

Under centering within organization parameterization, prior literature (e.g., Pituch and Stapleton 2012) has shown that we can obtain an estimate of the overall mediation effect (ME$_{211}$; Fig. 1c, d) as OME$_{211} = a(b_1 + b_2) = aB$. In terms of our example, the overall mediation effect quantifies the total improvement in practitioner use of a practice that accrues as a result of changes in both practitioner and collective attitudes toward this practice.

Power To track the statistical power with which we can discover the overall mediation effect, we can again draw on and extend the Monte Carlo interval test (Kelcey et al. 2017). Under maximum likelihood estimation of the parameters, we can sample from a multivariate normal distribution for the estimated path coefficients with the vector of means set to the anticipated values for the a and B path coefficients and the covariance matrix set to have diagonal terms equal to the corresponding path error variances (Kelcey et al. 2017). Similar to the 2-2-1 case, we track power in relation to hypothesized path coefficients in the planning stages by deriving the implied error variances as functions of the primary path coefficients (i.e., a, B, b_1, c'), the variance explained in the outcome at each level by covariates ($R^2_{Y^{L2}_{W,\overline{X}}}$ and $R^2_{Y^{L1}_{\overline{X},V}}$), the variance in the mediator explained by covariates at both levels ($R^2_{M^{L2}_{W,\overline{X}}}$ and $R^2_{M^{L1}_{\overline{X},V}}$), and the unconditional intracluster correlation coefficients of the outcome (ρ_Y) and mediator (ρ_M). We can then obtain draws from the posterior distributions and approximate the sampling distribution of the mediation effects by assembling the products of the draws. Power is estimated as the proportion of asymmetric intervals (e.g., 95%) that exclude zero. The net result is directly analogous to the 2-2-1 case—by specifying the mediation path coefficients, the variance explained by covariates, and the variance decompositions of the outcome and mediator, we can determine the sample size needed to achieve a specific power level.

Also similar to the 2-2-1 case, if we normalize the outcome and mediator to have unconditional means of zero and variances of one (i.e., $\tau^2_M + \sigma^2_M = \rho_M + (1-\rho_M) = 1$ and $\tau^2_Y + \sigma^2_Y = \rho_Y + (1-\rho_Y) = 1$), the organization-level mediator and outcome variances serve as the respective intracluster correlation coefficients (ρ_M and ρ_Y). Standardizing the outcome and mediator places the treatment-mediator coefficient on a standardized difference scale and the mediator–outcome path on a standardized regression coefficient scale. We can also apply the aforementioned alternative effect sizes. The 2-1-1 effect sizes become as follows: (a) single-cluster: $d_w = aB/\sigma_Y$; (b) multi-cluster or broad population: $d_T = aB/\left(\tau^2_Y\left(1-\frac{p(1-p)}{\tau^2_Y}(aB+c')^2\right)+\sigma^2_Y\right)$, and (c) cluster-level: $d_B = aB/\left(\tau^2_Y\left(1-\frac{p(1-p)}{\tau^2_Y}(aB+c')^2\right)\right)$.

Illustration Let us return to our working example to illustrate a power analysis concerning the detection of the overall mediation effect when a practitioner-level mediator is of interest. Consider the design of a multilevel mediation study that measures the extent to which an organization-wide training program (intervention) impacts practitioner use of motivational interviewing (outcome) by improving practitioner attitudes toward motivational interviewing (mediator). We focus on the total or overall mediation effect that captures the extent to which exposure to the training program produces changes in practitioner use of motivational interviewing by changing their individual and/or collective attitudes regarding the practice.

Let us continue with the design parameter values adopted in the previous illustration—assume we anticipate an intracluster correlation coefficient for the outcome of 0.30 (ρ_Y), that the covariates will explain 60% of the outcome and mediation variance at each level ($R^2_{Y^{L2}_{W,\overline{X}}} = R^2_{Y^{L1}_{\overline{X},V}} = R^2_{M^{L2}_{W,\overline{X}}} = R^2_{M^{L1}_{\overline{X},V}} = 0.60$), the direct effect of the treatment on the outcome is $c' = 0.10$, and that the treatment-mediator path coefficient is $a = 0.60$. Let us further adopt an intracluster correlation coefficient for the mediator of 0.30 (ρ_M), an overall mediator–outcome association of $B = 0.50$ with 0.10 owing to the practitioner-level association ($b_1 = 0.10$). If again 50% of the organizations are exposed to the intervention ($P = 0.50$) with 30 practitioners sampled at each organization (n_1), how many organizations must we sample to have a power level of 0.90?

To reach a power level of 90%, our analyses suggested just over 80 organizations would be required (approximately 40

treatment, 40 control). As a point of reference, about half as many (38) organizations would be required to detect the main or total effect of $ab + c' = c = 0.40$ (i.e., in a model excluding the mediator). We again conducted our analyses using *PowerUp-Mediator*. We provide a screenshot of the interface for 2-1-1 mediation designs in Fig. 2 (bottom). The software implements the Monte Carlo interval test as well as the Sobel and joint tests for mediation and provides a label and a brief description of the parameters needed to evaluate power. In contrast to the 2-2-1 case, the software conducts power analyses for the three types of mediation effects previously discussed—overall mediation (OME), individual or lower-level mediation (LME), and contextual or upper-level mediation (UME). Users input the anticipated values of the parameters that govern power and the software automatically estimates power under the joint and Sobel tests. Once users push the Run MC button, estimates of power under the Monte Carlo interval test are produced in about a minute depending on computing speed.

Lower-Level Mediation Effects

Under additional assumptions that render effects exchangeable or constant across practitioners and organizations, we can descriptively decompose the overall mediation effect into components that specifically flow through the practitioner- and organization-levels (Fig. 1b, c). As such, the second category we consider is the unique individual or lower-level mediation effect (LME). A lower-level mediation effect examines the extent to which the effects of an organization-level treatment on a practitioner-level outcome are transmitted through the practitioner-level component of the mediator (e.g., Krull and MacKinnon 2001; Pituch and Stapleton 2012). We can obtain an estimate of the lower-level mediation (LME) as $LME = ab_1$. With our running example, the lower-level mediation effect quantifies the improvement in practitioner use of motivational interviewing that accrues as a result of changes in individual attitude produced by the training when holding constant colleague attitudes (i.e., other practitioners at that organization). In assigning organizations to interventions, the lower-level mediation effect is not directly observable because practitioners at an organization experience the same intervention condition. However, if effects are approximately consistent across units, we can use practitioners at other organizations with a different intervention condition as reasonable proxy counterfactuals and descriptively identify the part of the overall mediation effect that owes specifically to changes at the individual-level.

Power For the lower-level mediation effect, the Monte Carlo interval test can be formed using an approach analogous to that of power for the overall mediation effect. We sample from a multivariate normal distribution with the vector of means set to the anticipated values for the a and b_1 path coefficients and the covariance matrix set to have diagonal terms equal to their corresponding path error variances. Identical to the overall mediation effect, the error variance of the individual-level mediation-outcome path coefficient ($\sigma_{b_1}^2$) can be tracked using the primary path coefficients (i.e., a, B, b_1, c'), the variance explained in the outcome at each level by covariates ($R^2_{Y_{W,\bar{X}}^{L2}}$ and $R^2_{Y_{\bar{X},V}^{L1}}$), the variance in the mediator explained by covariates at both levels ($R^2_{M_{W,\bar{X}}^{L2}}$ and $R^2_{M_{\bar{X},V}^{L1}}$), and the unconditional intracluster correlation coefficients of the outcome (ρ_Y) and mediator (ρ_M). Using these estimators, we can draw samples of the a and b_1 path coefficient from their respective normal distributions. In turn, the power of the Monte Carlo interval test can be obtained by drawing and multiplying samples of a and b_1 and recording the proportion of asymmetric confidence intervals that exclude no effect.

Illustration Let us now focus on the extent to which exposure to the training program improved practitioners' use of interviewing by operating specifically through individual attitudes (i.e., targeting lower-level mediation effects). In this particular example, we focus on differences among practitioners in the absolute sense—as opposed to the standing of practitioner relative to their organization means (Pituch and Stapleton 2012). Within this context, the raw or grand-mean centered model conceptually describes these lower-level mediation effects because it draws on the absolute standing of practitioners. However, as described earlier, such lower-level mediation effects can also be captured by the centering within organization formulation as long as we enter the organization means at the group-level. As a result, the formulas in this study can be directly applied to estimate power regardless of whether researchers intend to use no centering, centering across organizations, or centering within organizations in their analytic model.

Let us retain the parameter values previously outlined. Our power analysis now focuses on designs that can detect a lower-level mediation effect of $ab_1 = 0.06$. That is, we anticipate that 0.06 of the impact of the training on the outcome flows specifically through the practitioner-level mediator. If we adopt a type 1 error rate of 0.05, how many organizations must we sample to achieve 90% power to detect the lower-level mediation?

We again implemented the analyses in the *PowerUp-Mediator* software. User follows the same procedure outlined for power analyses for the overall mediation effect. The results suggest the required sample sizes—even for a relatively small mediation effect of 0.06—are much smaller for lower-level mediation than overall mediation. Even as few as 36 organizations (18 treatment, 18 control) would be sufficient in detecting a lower-level mediation effect as small as $ab_1 = 0.06$ with 90% power. More generally, when other factors are held

constant, lower-level mediation effects will typically have higher power than the overall mediation effect because the driving sample size is the number of practitioners rather than the number of organizations.

Upper-Level Mediation Effects

The final category we consider includes the mediation effect that represents the contextual association (Fig. 1b). The unique upper-level mediation effect focuses on the contextual or environmental effect of the mediator organization means on the outcome; that is, the association of the organization means beyond that which is supplied by the correlation between the outcome and practitioner-level mediator. The contextual or upper-level mediation effect estimates the increment in practitioner use of motivational interviewing that accumulates as a result of changes in colleagues' attitudes when holding constant individual practitioner attitudes. Again, this type of decomposition of the overall mediation effect requires the additional assumption that effects are approximately constant across practitioners and organizations.

Conceptually, the upper-level mediation effect is not directly obtained from the above formulation of the outcome model (and depicted in Fig. 1c) because the organization-mean parameterization orthogonalizes the practitioner- and organization-level mediator values such that the coefficients attached to the mediator organization means capture the sum of the practitioner- and organization-level outcome-mediator associations. More practically, the unique organization-level effects are represented by modifying the outcome model (4ab) so that the practitioner-level mediator values reflect the absolute standing (e.g., grand-mean centered) or original mediator values of practitioners within organizations rather than the organization-mean centered values (Fig. 1b; Pituch and Stapleton 2012). That is, the practitioner-level mediator enters the outcome model as

$$Y_{ij} = \beta_{0j} + b_1 M_{ij} + \beta_1\left(X_{ij} - \overline{X}_j\right) + \beta_2 V_{ij} + \varepsilon_{ij}^Y \qquad \varepsilon_{ij}^Y \sim N\left(0, \sigma_{Y|}^2\right) \text{(5a)}$$

$$\beta_{0j} = \gamma_{00} + b_2 \overline{M}_j + c' T_j + \gamma_{01} W_j + \gamma_{02} \overline{X}_j$$
$$+ u_{0j}^Y \qquad u_{0j}^Y \sim N\left(0, \tau_{Y|}^2\right) \qquad \text{(5b)}$$

In this uncentered formulation (or grand-mean centered approach), the coefficient attached to the mean mediator (b_2) now captures the contextual or unique organization-level conditional association with the outcome because the organization and individual mediator values are no longer orthogonal. In terms of the treatment-mediator path, because the treatment is assigned to organizations we cannot disentangle its effects on the individual- and organization-level components of the mediator even under standard additional assumptions.

Alternatively, we can retain the centering within organizations (Eq. 4ab; Fig. 1c) and estimate the contextual or upper-level mediation effect as the differences between the overall and lower-level mediation effects. For the upper-level mediation effect (UME), this yields UME $= ab_2 = a(B - b_1)$.

Power For the upper-level mediation effect, the Monte Carlo interval tests can be adapted in a straightforward manner to obtain a test of upper-level mediation. We sample from a multivariate normal distribution with the vector of means set to the anticipated values for the a and b_2 path coefficients and the covariance matrix set to have diagonal terms equal to the corresponding path error variances. Because the b_2 path can be represented using the difference between the overall and lower-level mediator–outcome association, we can estimate its error variance as a function of the previously outlined error variances and the exact same parameters. Power for upper-level mediation follows the same structure as prior tests.

Illustration Let us target the extent to which exposure to the training program improved practitioners' use of interviewing by operating specifically through contextual changes in organization-wide attitudes (i.e., targeting upper-level mediation effects). That is to say, suppose we were precisely interested in how training-induced changes in organization-level attitudes impacted individual practitioner use of motivational interviewing. Maintaining the same parameter values noted above, we focus on designs that can detect an upper-level mediation path of $ab_2 = 0.24$. Specifically, we continue to anticipate that the overall mediator–outcome association will be $B = 0.5$, the practitioner-level component of the mediator–outcome association will be $b_1 = 0.1$, and the contextual component of the mediator–outcome association will be $b_2 = B–b_1 = 0.4$.

Once again, we used the *PowerUp-Mediator* software and the aforementioned procedure. In contrast to lower-level mediation, the results suggest that the required sample sizes to achieve a certain power level will tend to be larger than those necessary to detect the overall or lower-level mediation effects. In this instance, the results indicate that assigning approximately 130 organizations (65 treatment, 65 control) would yield a 90% chance of discovering intervention-induced changes in interviewing practice operating through the collective or average attitudes of practitioners at organizations.

Discussion

Conducting investigations with the capacity to test a more comprehensive set of effects—such as mediation effects—has become a prominent aim of prevention science research (e.g., Gottfredson et al. 2015). Investigations that probe a clearly articulated theory of an intervention help to disentangle

the black box of its effects by assessing an action theory that delineates how the intervention affects intermediate variables and a conceptual theory that outlines how those intermediate variables pass on the intervention effects to the outcome (Gottfredson et al. 2015). A key to ensuring that researchers can select sufficiently powerful designs regarding these types of hypotheses is delineating the roles of parameters that govern such power.

In this study, we outlined several complementary types of mediation effects designed to unpack the potential pathways researchers often consider and outlined a powerful resampling test of multilevel mediation whose behavior can be easily tracked and understood even before data collection. The power of the Monte Carlo interval test can be easily tracked using the anticipated magnitudes of the path coefficients, the intracluster correlation coefficients, and the variance explained in the outcome and mediator by covariates. To make these power analyses more accessible to researchers, we have developed and illustrated the analyses in a free Excel-based program *PowerUp-Mediator*. We have also developed similar routines in the *R* package *PowerUpR*. Both programs use a simple interface where users input the anticipated parameter values and receive as the output power estimates under different tests.

Compliance with Ethical Standards

Conflict of Interest The authors declare that they have no conflicts of interest.

Ethical Approval This article does not contain any studies with human participants or animals performed by any of the authors.

Informed Consent The manuscript does not report any studies with human participants or animals; no study was conducted requiring informed consent.

References

Cox, K. & Kelcey, B. (2018). Optimal Sample Allocation in Group-randomized Studies of Multilevel Mediation with a Group-level Mediator. *Journal of Experimental Education* (in press).

Dong, N., Kelcey, B., Spybrook, J., & Maynard, R. A. (2016). PowerUp!-Mediator: A tool for calculating statistical power for causally-defined mediation in cluster randomized trials. (Version 0.4) [Software]. Available from http://www.causalevaluation.org/. Accessed 1 May 2007.

Gottfredson, D. C., Cook, T. D., Gardner, F. E., Gorman-Smith, D., Howe, G. W., Sandler, I. N., & Zafft, K. M. (2015). Standards of evidence for efficacy, effectiveness, and scale-up research in prevention science: Next generation. *Prevention Science, 16*, 893–926.

Kelcey, B., & Phelps, G. (2013). Considerations for Designing Group Randomized Trials of Professional Development with Teacher Knowledge Outcomes. *Educational Evaluation and Policy Analysis, 35*, 370–390.

Kelcey, B., Dong, N., Spybrook, J., & Cox, K. (2017). Statistical power for causally-defined indirect effects in group-randomized trials. *Journal of Educational and Behavioral Statistics.*

Kelcey, B., Dong, N., Spybrook, J., & Shen, Z. (2017a). Experimental Power for Indirect Effects in Group-randomized Studies with Group-level Mediators. *Multivariate Behavioral Research, 52*(6), 699–719.

Kelcey, B., Dong, N., Spybrook, J. & Cox, K. (2017b). Statistical power for causally-defined indirect effects in group-randomized trials. *Journal of Educational and Behavioral Statistics.*

Kelcey, B. & Shen, Z. (2018). Effective and Efficient Experimental Design 2-1-1 Mediation Studies. *Journal of Experimental Education* (in press).

Krull, J. L., & MacKinnon, D. P. (2001). Multilevel modeling of individual and cluster level mediated effects. *Multivariate Behavioral Research, 36*, 249–277.

Phelps, G., Kelcey, B., Liu, S., & Jones, N. (2016). Informing Estimates of Program Effects for Studies of Mathematics Professional Development Using Teacher Content Knowledge Outcomes. *Evaluation Review, 40*, 383–409.

Pituch, K. A., & Stapleton, L. M. (2012). Distinguishing between cross- and cluster-level mediation processes in the cluster randomized trial. *Sociological Methods & Research, 41*, 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.

Texas Christian University [TCU] (2005) Institute of Behavioral Research. The Organizational Readiness for Change: Treatment Staff Version (TCU ORC-S). Available at: www.ibr.tcu.edu.

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

Schwartz, R. (2010). Motivational interviewing (patient-centered counseling) to address childhood obesity. *Pediatric Annals, 39*(3), 154–158.

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.

Stapleton, L., Pituch, K., & Dion, E. (2015). Standardized effect size measures for mediation analysis in clusterrandomized trials. *Journal of Experimental Education, 83*(4), 547–582.

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

Williams, N., & Glisson, C. (2014). Testing a theory or organizational culture, climate and youth outcomes in child welfare systems: A United States national study. *Child Abuse and Neglect, 38*, 4.

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