Abstract Title Page

Title:
Degenerate Power in Multilevel Mediation: The Non-monotonic Relationship Between Power & Effect Size

Authors and Affiliations:
Ben Kelcey
University of Cincinnati

Abstract Body

Background / Context:

There is a long established tradition of investigating the mechanisms through which a treatment is presumed to impact an outcome using mediation analyses (e.g., Baron & Kenny, 1986). Mediation analyses examine the extent to which a treatment has an indirect effect on an outcome by examining how changes in a mediator produced by exposure to a treatment manifest as changes in an outcome. Like studies focused on the detection of a main or total effect, a principal consideration in the design of studies examining mediation is the power with which mediation effects can be detected if they exist (e.g., Raudenbush, 1997). However, unlike studies concerning the detection of main effects, literature has not established power formulas for detecting mediation effects in multilevel designs and described precisely how changes in parameter values can influence and even undermine a study's power to detect mediation effects.

Purpose / Objective / Research Question / Focus of Study:

The purpose of this study was twofold. First, we derived power formulas for multilevel mediation effects. Second, we developed formulas that delineate the conditions under which power is maximized and the rate with which it declines as a function of the magnitude of path coefficients and used these results to investigate the behavior of power as a function of design parameters. The results indicated that unlike the power to detect total effects, the power to detect mediation effects is *not* a monotonic function of effect size but rather a complex function governed by the decomposition of the total effect (see for example Figure 1).

Significance / Novelty of study:

Despite the widespread use of mediation analyses and the recent emphasis on well-planned studies, literature is incomplete in identifying and delineating considerations for designing studies adequately powered to detect multilevel mediation effects. In particular, literature is incomplete in that (a) there are no available formulas detailing the power to detect multilevel mediation effects and (b) there are no available expressions describing the conditions under which (i) power is maximized given various parameter constraints and (ii) the conditions under which power becomes a non-monotonic function of effect size such that power actually declines as effect size increases. In this study, we developed a framework to help researchers design multilevel mediation studies by deriving formulas to assess the power of a design, describe the complex and atypical behavior of power in studies of mediation, and delineate the conditions under which it is maximized for a given set of parameter values.

Statistical, Measurement, or Econometric Model:

For our first objective concerning the development of power formulas for multilevel mediation effects, we outline our derivations by describing the Sobel mediation test for a 2-2-1 multilevel mediation model but note that the proposed approach extends to other inferential tests and designs. Consider a study where students are nested within schools and its hypothesized that a school level (level 2) treatment impacts a school level (level 2) mediator which impacts a student (level 1) outcome. To assess the extent to which a school level mediator mediates the relationship between the treatment program and student achievement, we draw on a multilevel linear structural equation framework using the following set of equations

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

Outcome model (Level 1) $$Y_{ij} = \beta_{0j} + \beta_{1j}(X_{ij} - \overline{X}_j) + \varepsilon_{ij}^Y \qquad \varepsilon_{ij}^Y \sim N(0, \sigma_{Y|}^2) \qquad (2a)$$

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

In the mediation equation (1), we use M_j as the mediator for school *j*, *W* as school level covariates, $\bar{X}$ as school level aggregates of student level covariates, with $\pi_.$ as the respective regression coefficients, T_j as the treatment assignment with coefficient *a*, and ε_j^M as the error term. In the outcome equations (2ab), we use a hierarchical linear model such that Y_{ij} is the outcome for student *i* in school *j*, X_{ij} is student level covariates with coefficients $\beta_.$, and ε_{ij}^Y is the level one error term. At the school level, we introduce $\gamma_.$ as the respective regression coefficients, *b* as the conditional relationship between the mediator and the outcome, c' as the direct effect of the treatment, and u_{0j} as the school level random effects.

One estimate of the mediation effect using this system of equations is the product of the *a* and *b* coefficients (Zhang, Zyphur, & Preacher, 2009). The asymptotic variance of the mediation effect can then be estimated using the Delta method (Sobel, 1982)

$$\sigma_{ab}^2 = b^2\sigma_a^2 + a^2\sigma_b^2 \tag{3}$$

where σ_{ab}^2 is the variance of the mediation effect, σ_a^2 is the error variance of the *a* coefficient in equation 2.1, and σ_b^2 is the error variance of the *b* coefficient in Equation 2.2b. In turn, the statistical significance of the mediation effect can be approximated using the Sobel test

$$z_m = ab / \sqrt{b^2\sigma_a^2 + a^2\sigma_b^2} \tag{4}$$

where z_m is approximately normally distributed in large samples (Sobel, 1982).

In our study, we extend these derivations to develop a closed form expression for the Sobel test statistic for multilevel mediation. Based on derivations omitted for space, we show that with standardized variables, the Sobel test statistic for multilevel mediation is

$$z_m = \frac{ab}{\sqrt{b^2\left(\frac{(1-R_M^2)}{d_M(1-R_T^2)}\right) + a^2\left(\frac{((1-R_{Y^{L2}}^2)\rho + (1-R_{Y^{L1}}^2)(1-\rho)/n)}{d_Y(1-R_M^2)}\right)}} \tag{5}$$

where ρ is the intraclass correlation coefficient, *n* is the within group sample size, $d_.$ are the degrees of freedom for the mediator (*M)* and outcome (*Y*) models, R_M^2 is the proportion of variance in the mediator explained by the treatment and covariates, R_T^2 is the proportion of variance explained in the treatment by covariates (e.g., $R_T^2 = 0$ under random assignment), and $R_{Y^{L2}}^2$ and $R_{Y^{L1}}^2$ are the proportion of school and student level outcome variance explained by all variables. Assuming the alternative hypothesis is true, the test statistic in Equation 5 follows a non-central distribution with (5) as the non-centrality parameter. The power of a two-sided test to detect a multilevel mediation effect is

$$P(z_m > z_{critical}) = 1 - \Phi(z_{critical} - z_m) + \Phi(-z_{critical} - z_m) \tag{6}$$

where Φ is the normal distribution and $z_{critical}$ is the chosen critical value (e.g., 1.96). Using this formula, researchers can now assess the power of a proposed study to ensure the design can detect mediation effects with a sufficiently large probability if they exist.

For our second objective concerning the conditions under which power is maximized, we continue with the aforementioned Sobel test for a 2-2-1 multilevel mediation model but drop the covariates. However, we note that the proposed approach again extends to models which incorporate covariates and other inferential tests. We begin by reformulating the test statistic in equation (5) in terms of path coefficients

$$z_m = \frac{ab}{\sqrt{\left(\frac{b^2 d_Y - 2a^2 b^2 d_Y + a^4 b^2 d_Y + (\rho + (1-\rho)/n)a^2 d_m - a^2 b^2 d_m - a^2 c'^2 d_m - 2a^3 bc' d_m}{d_m d_Y (1-a^2)}\right)}} \tag{7}$$

where notation is unchanged. Analysis of this function reveals that power is a non-monotonic function of effect size because the error variance of the mediation effect is inflated proportional to the magnitude of the *a* path. More conceptually, although the magnitude of the mediation effect (*ab*) increases as the *a* path increases, an increase in *a* also amplifies the error variance of the mediation effect (σ^2_{ab}) because it induces collinearity between the mediator and treatment in the outcome equation (2ab). As a result, holding other factors constant, increases in the magnitude of the *a* path can serve to increase *or* decrease power.

For this reason, we developed expressions to identify the conditions under which constrained maximum power is obtained and describe how changes in parameters impact levels of power for a given set of constraints. Because the magnitude of the *a* path can serve to increase or decrease power, we describe these conditions using the first derivative of equation (7) in terms of *a*,

$$\frac{\partial z_m}{\partial a} = \frac{(a^2b^2d^2 - a^4b^2d^2)(-6ab^2d + 4a^3b^2d - 6a^2bc'd - 2ac'^2d + 2ad(\rho + \frac{1-\rho}{n}))}{(b^2d - 3a^2b^2d + a^4b^2d - 2a^3bc'd - a^2c'^2d + a^2d(\rho + \frac{1-\rho}{n}))^2} + \frac{2ab^2d^2 - 4a^3b^2d^2}{b^2d - 3a^2b^2d + a^4b^2d - 2a^3bc'd - a^2c'^2d + a^2d(\rho + \frac{1-\rho}{n})} \tag{8}$$

where maximum power then occurs when

$$b^2n - 2b^2na^2 + bc'na^3 + (-1 + 2b^2n + c'^2n + \rho - n\rho)a^4 + bc'na^5 = 0 \tag{9}$$

In turn, we use this result along with the power formula derived for objective one to (a) estimate power, (b) understand the non-monotonic nature of the power curve in relation to the magnitude of the mediation effect, and (c) design efficient studies of multilevel mediation.

Findings / Results:

Based on the derivations above, we briefly highlight two conceptual results. First, the power to detect a mediation effect is a non-monotonic function of the magnitude of the *a* path but a strictly increasing function of the magnitude of the *b* and c' paths and a strictly decreasing function of the intraclass correlation coefficient (see Figure 1 and illustration below for an example). Both the *a* and *b* paths serve to increase the mediation effect (*ab*) and ultimately increase the magnitude of the numerator in equation (7). However, the *a* path also serves as a variation inflation factor because it necessarily introduces collinearity between the mediator and treatment in the outcome equation (2ab), thus increasing error variance as captured by the denominator in equation (7). Taken as a whole, the uncertainty introduced by this collinearity eventually overtakes the increase in the mediation effect supplied by the *a* path, resulting in a net loss of power. With regards to the intraclass correlation coefficient, like power functions for total effects, increases in the intraclass correlation coefficient simply operate by amplifying the error variance thereby reducing power. In contrast, the c' path serves to reduce the error variance of the mediation effect because its magnitude captures covariance between the outcome and treatment that is unrelated to the mediator, resulting in an increase in power.

Second, the location of the inflection point of the power curve does not depend on the school level sample size (see equation 9 and Figure 2). However, the maximum attainable power and the rate with which power decreases are proportional to school sample size. As a result, the scale and non-monotonic nature of power is practically relevant only when school sample sizes are less than say 150 and/or when a study anticipates the *a* path to be large relative to the *b* path.

Usefulness / Applicability of Method:

Consider an example in which intact schools are randomly assigned to participate in a whole school reform program or business as usual. A critical feature of whole school reform programs is that they mobilize educators to coordinate teaching across subjects to achieve common aims (Raudenbush, 2003). Literature has suggested that a critical mediator of the impact of whole school reform on student achievement is school climate (e.g., MacNeil, Prater, & Busch, 2009). In theory, participation in a whole school reform program (treatment) should improve the school climate (mediator) which in turn improves achievement (outcome).

Let us assume that we intend to design a study of 50 schools with 20 students per school with an intraclass correlation coefficient of 0.20 and assume researchers anticipate a *b* path magnitude of about 0.25 with complete mediation. Based on the results of our derivations, we should conduct power analyses for a range of coefficients to identify reasonable study designs that are robust to reasonable variations in these (and other) parameters. For instance, we can begin by plotting the power as a function of the magnitude of the mediation effect in terms of the *a* and *b* paths (see Figure 1). Our range of the *b* coefficients might include *b*=0.15, 0.25, 0.35. Evident from Figure 1, when the magnitude of the *b* path is 0.25, our design will have power at the nominal level of 0.8 only when the *a* path is 0.61. If the *b* path is somewhat smaller than anticipated, say 0.2 or 0.15, the power to detect multilevel mediation will never exceed 0.6 or 0.4, respectively, for any value of the *a* path. That is, with an expected *b* path of 0.25, the proposed design will almost never achieve .8 power regardless of the magnitude of the *a* path. In this way, the proposed design would seem to be sensitive to reasonable variations in design parameters such that we should consider modifying design parameters, e.g., larger sample sizes.

Conclusions:

Our findings suggest that the power to detect mediation is more complex than the power to detect total effects. Perhaps most notably is that power is a non-monotonic function of the magnitude of the *a* path. Although potentially counterintuitive based of the monotonic nature of power curves typically associated with total effects, consideration of collinearity issues quickly illuminates this discrepancy and delineates the basis for this non-monotonic relationship.

The complex nature of the power curve suggests that researchers investigating mediation need to carefully consider the far reaching effects of study features and design choices. For instance, the results of our analysis indicated that when designing a study of multilevel mediation, one must consider the *individual* magnitudes of the *a* and *b* paths and not just the magnitude of their product. In this way, researchers will need to either heavily anchor the expected magnitudes of these individual paths in previous empirical research and/or consider sample sizes that are sufficiently large to withstand the potential for variance inflation effects. The results also imply that designs which call upon mediators that are proximal to the outcome (i.e., *b* is relatively large) will be better suited to detect mediation effects than designs which call for mediators more proximal to the treatment (i.e., *a* path is large relative to the *b* path). Similarly, the results suggest that, when possible, partially mediated designs may be desirable because they tend to reduce the variance of the estimated mediation effect.

Although the study presented focused on inferences drawn from the Sobel test for multilevel mediation, the results extend to other designs and tests including single level designs and bootstrapped confidence intervals. In this way, the findings from this study expand the scope and quality of designs for studies of mediation because they equip researchers with the tools to consider how study features and design choices impact the probability with which they can detect effects.

Appendices

Appendix A. References

Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic and statistical considerations. *Journal of Personality and Social Psychology, 51*, 1173-1182.

MacNeil, A. J., Prater, D. L., & Busch, S. (2009). The effects of school culture and climate on student achievement. *International Journal of Leadership in Education*, *12*(1), 73–84.

Raudenbush, S. W. (1997). Statistical analysis and optimal design for cluster randomized trials. *Psychological Methods*, *2*(2), 173–185.

Raudenbush, S. W. (2003, October). Designing Field Trials of Educational Innovations. In *a national invitational conference, Conceptualizing Scale-UP: Multidisciplinary Perspectives, Washington, DC*. Washington, DC.

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

Zhang, Z., Zyphur, M. J., & Preacher, K. J. (2009). Testing Multilevel Mediation Using Hierarchical Linear Models: Problems and Solutions. *Organizational Research Methods*, *12*(4), 695–719.

Appendix B. Tables and Figures

Figure 1

Power as a function of the magnitude of the *a* and *b* paths

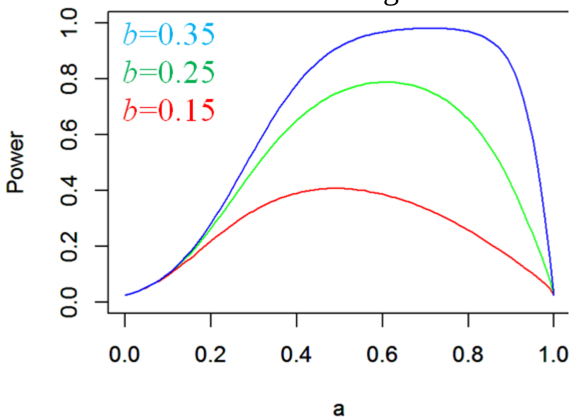

Figure 2:Power as a function of the magnitude of the *a* path and level two sample size