Published in final edited form as:

J Exp Educ. 2014 January 1; 82(3): 283–306. doi:10.1080/00220973.2013.813360.

Tests of Mediation: Paradoxical Decline in Statistical Power as a Function of Mediator Collinearity

T. Mark Beasley

University of Alabama at Birmingham

Abstract

Increasing the correlation between the independent variable and the mediator (a coefficient) increases the effect size (ab) for mediation analysis; however, increasing a by definition increases collinearity in mediation models. As a result, the standard error of product tests increase. The variance inflation due to increases in a at some point outweighs the increase of the effect size (ab) and results in a loss of statistical power. This phenomenon also occurs with nonparametric bootstrapping approaches because the variance of the bootstrap distribution of ab approximates the variance expected from normal theory. Both variances increase dramatically when a exceeds the b coefficient, thus explaining the power decline with increases in a . Implications for statistical analysis and applied researchers are discussed.

Keywords

Mediation Analysis; Power; Collinearity; Tolerance; Bootstrap

INTRODUCTION

Relations between variables are often more complex than simple bivariate relations between a predictor and an outcome criterion. Rather these relations may be modified by the addition of a third variable in the research design (Fairchild & MacKinnon, 2009). Statistical mediation analysis is used to determine whether the effect of a causal agent on an outcome variable is transmitted through a third intermediate variable (see Figure 1). Mediation analysis has a long history (MacCorquodale & Meehl, 1948) and has been a topic of interest to applied researchers in a variety of social science disciplines (e.g., communications; psychology; sociology; program evaluation; educational research).

For many applications in the social sciences, the utility of mediation analysis is based on theories for explaining the processes underlying intervention effects or other potential causal agents. *Conceptual Theory* focuses on how potential mediators are related to outcome variables and provides justification for targeting certain intermediate variables. Typically, once potential mediators are identified, *Action Theory* is used to develop programs or interventions to modify the intermediate variables (Chen, 1990, Lipsey, 1993; MacKinnon, 2008). For example, Varnhagen and Zumbo (1990) found that student learning of statistical concepts was difficult to affect directly through instruction; however, the educational format (i.e., instructional treatment) influenced a mediator (i.e., student attitude) and had an indirect effect on the outcome (i.e., student performance) through a mediational process. Although it

has not received as much attention outside of the social sciences, mediation analysis is of growing interest in the other fields such as psychiatry (Meyer-Lindenberg & Weinberger, 2006), neurobiology (Stoltenberg et al. 2002), behavior genetics (Munafò, 2006), neuropsychiatry (Wilkins, Girard, & Cheyne, 2012), experimental gerontology (Barger, Kayo, Pugh, Prolla, & Weindruch, 2008), gene expression studies (Allison, Heo, Schork, Wong, & Elston, 2002), health behavior (Veitch et al., 2012), nutrition (Hanks et al., 2010; Forestell & Mennella, 2012), ophthalmology (Popescu et al., 2012), pain management (Meulders, Vansteenwegen, & Vlaeyen, 2012) and quality-of-life research (Cox et al., 2011;).

Numerous statistical methods for testing mediation effects have been proposed with no absolute consensus on which is best (Albert, 2008; MacKinnon, Lockwood, Hoffman, West, & Sheets, 2002); however, the most commonly used methods involve testing the product (ab) of the a coefficient that represents the effect of the Independent Variable (X) on the Mediator (Z) and the b coefficient for the partial relationship of the Outcome (Y) with the mediator (see Figure 1). In both simulation studies and research applications of mediation analysis, investigators have noted that the power to detect the mediated effect can decline as the effect size (ab) increases. Other studies (e.g., Fritz Taylor, & MacKinnon, 2012) have discussed this power decline as an unexpected anomaly with the magnitude of b being the most important factor in explaining this phenomenon. In this paper, this decline in statistical power is presented as a function of the a coefficient, which by definition is the collinearity between X and Z . Increases in collinearity (a coefficient) result in concomitant reductions in the tolerance of the mediator, which inflate the variance of the b coefficient. The variance inflation of the b coefficient due to increases in a also inflates the standard error of the product and at some point this increase in variance outweighs the increase of the effect size (ab), resulting in a loss of statistical power for testing mediation. A simulation study demonstrates that nonparametric bootstrap methods that do not directly estimate standard errors are not immune to the decline in statistical power because the variance of bootstrap distributions for ab closely approximates the variance expected from normal-theory tests and thus are affected by the magnitude of a in the same manner as normal-theory tests. A concrete, computational example is provided. Implications for researchers conducting statistical analyses, performing power and sample size calculations, and designing studies to investigate mediational processes are discussed.

TESTS OF MEDIATION

One of the most popular approaches for assessing mediation is to fit a sequence of three linear regression models (Baron & Kenny, 1986). Let Y_i denote the outcome for individual i . Z_i denotes a potential mediator variable and X_i is a potential causal factor (i.e., antecedent, treatment, intervention, or independent variable) of interest. The three linear regression models for mediation analysis can be expressed in standardized notation (i.e., no intercepts) as:

$$\text{Model 1: } Y_i = \tau X_i + \varepsilon_{1i} \quad (1)$$

$$\text{Model 2: } Z_i = \alpha X_i + \varepsilon_{2i} \quad (2)$$

$$\text{Model 3: } Y_i = \beta Z_i + \tau' X_i + \varepsilon_{3i} \quad (3)$$

where τ , a , β , and τ' represent population regression parameters estimated by sample coefficients t , a , b , and t' , respectively. Standard linear model assumptions that the errors (ε 's) are independent and normally distributed with a mean of zero and constant variance apply. The relationship between an independent variable (X) and the outcome (Y) is decomposed into direct and indirect (mediated) effects (see Figure 1). The amount of mediation can be estimated in two ways: the difference between two sample regression coefficients [$t - t'$] or the product of two sample regression coefficients [ab]. MacKinnon, Warsi, and Dwyer (1995) demonstrated that in many circumstances these two sample estimates of mediation are identical (i.e., $ab = [t - t']$), regardless of whether multiple linear regression or path analysis is used.

For the “causal steps” approach, the intermediate variable Z is considered to mediate the effect of X on Y if: (1) $H_{0(1)}$: $\tau = 0$ in Model 1 is rejected; (2) $H_{0(2)}$: $\alpha = 0$ in Model 2 is rejected; and (3) $H_{0(3)}$: $\beta = 0$ in Model 3 is rejected. If all three of these null hypotheses are rejected and $H_{0(4)}$: $\tau' = 0$ in Model 3 is also rejected, then the results indicate that variable Z *partially* mediates the X - Y relationship (MacKinnon, 2008). To demonstrate *complete mediation*, the Baron and Kenny (1986) “four rule” approach requires rejecting the first three null hypotheses and not rejecting the fourth.

Other methods assess mediation by performing a single test but do not require the X - Y partial relationship (t') to be non-significant (MacKinnon et al., 1995). McGuigan & Langholtz (1988) proposed a t -statistic for testing H_0 : $[\tau - \tau'] = 0$; however, the performance of this procedure has been inconsistent in simulation studies and MacKinnon et al. (1995) concluded that it should not be used when the independent variable (X) is binary. Product tests use ab as the estimate of the mediation effect and perform a single test of H_0 : $a\beta = 0$. Sobel (1982) derived a first-order standard error for the product as a weighted composite of the standard errors for the a and b coefficients, $SE^2_{(ab)S} = b^2 SE^2_{(a)} + a^2 SE^2_{(b)}$. The derivation of the Sobel (1982) standard error assumes that a and b are independent, which is true when linear model assumptions are met (MacKinnon, 2008). Other approaches include the covariance of the a and b coefficients, which is typically small. Aroian (1947) adds this covariance term, $SE^2_{(ab)A} = b^2 SE^2_{(a)} + a^2 SE^2_{(b)} + SE^2_{(a)} SE^2_{(b)}$. Goodman (1960) subtracts this covariance term, $SE^2_{(ab)G} = b^2 SE^2_{(a)} + a^2 SE^2_{(b)} - SE^2_{(a)} SE^2_{(b)}$, to form an unbiased estimate of the variance of the mediated effect; however, $SE^2_{(ab)G}$ can yield negative variance estimates. All three of these approaches use the unit normal distribution to obtain p -values for z -statistics (i.e., $z = ab / SE_{(ab)}$) or critical values to form symmetric confidence intervals (CIs).

In general, the product of two random variables is not normally distributed, even if both components are normally distributed (Lomnicki, 1967; Springer & Thompson, 1966). Therefore, there are concerns over the accuracy of normal-theory tests and CIs for the ab product because of expected departures from the normality assumption. In order to deal with

the non-normality of the product distribution, MacKinnon, Fritz, Williams, & Lockwood (2007) developed a FORTRAN program called PRODCLIN that uses numerical integration (NI) to estimate the distribution of the product of two normally distributed variables to construct asymmetric CIs; however if the normality assumptions underlying Models 2 and 3 are violated this procedure may have questionable statistical properties (Biesanz, Falk, & Savalei, 2010).

Bootstrapping has become an increasingly popular approach to assessing mediation effects due to concerns over violating normality assumptions (Shrout & Bolger, 2002; Hayes, 2009; Preacher & Hayes, 2008). Bootstrapping is a general approach to statistical inference based on randomly resampling from the observed data (Efron & Tibshirani, 1993). Bootstrap distributions are expected to follow the properties of a statistic if random samples were repeatedly drawn from the same population, and they have been shown to provide good approximations to the theoretical sampling distribution for many statistics (Zieffler, Harring, & Long, 2011).

Applied to mediation analysis, bootstrapping creates an empirical sampling distribution for ab (Bollen & Stine, 1990; MacKinnon et al., 2004). Bootstrap samples are the same size as the original sample (N) and entire row vectors of the data (i.e., Y , Z , and X) are randomly resampled *with replacement* N times. Therefore, any given bootstrap sample may have the data for any case appear more than once or not at all. Once a bootstrap sample is selected Models 1 through 3 are performed and a parameter estimate (ab) is computed. This process is repeated B times and the mean or the median of the B ab values is used as the point estimate.

Nonparametric bootstrapping approaches do not estimate the standard error (Efron & Tibshirani, 1998). Rather, the the observed data is used as a proxy for the population and the distribution of the B bootstrapped estimates of ab is used as the sampling distribution in order to empirically estimate CIs. The percentile bootstrap (PB) estimates the $100(1-\alpha/2)$ and $100(\alpha/2)$ percentiles of the bootstrap distribution of ab as the upper and lower bounds of the CI, where α is the significance level. For example, with a significance level of $\alpha = 0.05$ and $B = 5000$ bootstrap samples, the 97.5th percentile is estimated as the 4875th largest value in the bootstrap distribution of ab estimates. That is, only 125 (2.5%) values are larger than the 4875th ranked value. Likewise, 2.5th percentile is estimated as the 126th largest value of ab in the bootstrap distribution (i.e., only 125 (2.5%) values are smaller). If the bootstrap CI does not contain zero, then the null hypothesis of no mediation ($H_0: a\beta = 0$) is rejected. Efron and Tibshirani (1993) proposed bias-corrected (BC) and accelerated bias-corrected (BCA) bootstrap approaches to improve the percentile bootstrap CIs by adjusting the percentile positions for the bias between the sample and the bootstrap point estimates and variance inflation (i.e., acceleration) due to asymmetry and bias. These approaches to bootstrap estimation have been endorsed by many researchers and incorporated into the INDIRECT software for mediation analysis (Preacher & Hayes, 2008).

Statistical Power for Tests of Mediation

Statistical power for mediation analysis is a complicated function of the sample size, the significance level, the presumed strength of the population associations of the mediator with

the independent variable (α parameter in Model 2) and the outcome (β parameter in Model 3), and the distribution of their product. Fritz and MacKinnon (2007) have tabled estimates of sample size needed to obtain 80% power for several methods and effect sizes at an $\alpha = 0.05$ significance level. The Baron and Kenny “four rule” method has consistently demonstrated the least statistical power of any approach to assessing mediation (e.g., Fritz & MacKinnon, 2007; Fritz et al. 2012). Of the normal-theory tests, the Sobel test has performed best in simulation studies (e.g., MacKinnon et al., 1995), and it typically maintains valid Type I error rates with sample sizes of $N > 50$. The Numerical Integration (NI) approach generally maintains valid Type I error rates and has greater power than normal-theory tests. The nonparametric Percentile Bootstrap (PB) performs similarly to the NI approach when the linear model assumptions underlying Models 2 and 3 hold (Biesanz et al., 2010; Fritz & MacKinnon, 2007; Fritz et al., 2012).

Although several simulation studies have examined the validity, accuracy, and power of these various mediation tests (e.g., Cheung, 2007, 2009; Fritz & MacKinnon, 2007; MacKinnon et al., 2004; MacKinnon, Lockwood et al., 2002; Preacher & Hayes, 2008; Williams & MacKinnon, 2008), all methods have also produced curious findings. In terms of bootstrapping, simulation studies have shown that the BC and BCA bootstrap tests of mediation have more statistical power than NI, PB, and the Sobel test; however, BC and BCA also tend to inflate Type I error rates, regardless of number of bootstrap samples (Cheung, 2007; MacKinnon et al., 2004). Thus, their enhanced power may simply be due to elevated false positive rates (Fritz et al., 2011).

A second paradoxical result is that for all methods the power to detect the mediated effect can decline as effect size (ab) increases. Specifically, this reduction in power can occur when the a coefficient becomes substantially larger than the b coefficient. Intuitively, it would seem that the power of product tests would increase with $|a|$, because the numerator, ab , increases in absolute value. A few authors (Beasley, 2012; Fritz et al., 2012) have noted that increases in $|a|$ result in increases in $SE^2_{(b)}$ in the denominator of the product tests. Hoyle and Kenny (1999) briefly discuss the power of testing $H_{0(3)}: \beta = 0$ and $H_{0(4)}: \tau' = 0$ in Model 3 as a function of the collinearity of X and Z . They state (p. 201) that the degree a affects the power of product tests is “less clear because both a and b contribute to the product.” Fritz et al. (2012) concluded that the magnitude of b was the most important factor in the power functions. Although Fritz et al. (2012) manipulated the a coefficient in their simulations, it was not included in their models to predict power. For the purposes of their study, they aggregated power results across levels of a to calculate what they termed “power stagnation.” Therefore, the decline in statistical power was not presented as a function of the a coefficient. Furthermore, these authors did not discuss this issue explicitly in terms of variance inflation of the product as a function of the a coefficient.

Standard Errors of the Sobel Test

Due to reported power advantages and the fact that they do not rely on normal theory assumptions, bootstrap methods have become an increasingly popular method for testing mediation (Hayes, 2009). Yet, the Sobel normal theory test remains one of the most widely-used tests in mediation analysis because of its accessibility, ease of computation, and

generally robust performance (MacKinnon et al., 1995; Fritz et al., 2012). Furthermore, for the purposes of this paper, the standard error of the Sobel test provides a vehicle for demonstrating variance inflation of the product and the resultant power decline of the product tests as a function of mediator collinearity (a coefficient).

Regression analyses can be expressed in terms of standardized regression coefficients and bivariate correlations (e.g., Pedhazur, 1997). The explained variation for Model 1 is: $R^2_1 = r^2_{YX} = t^2 = (t' + ab)^2$ and the squared standard error for the t coefficient is: $SE^2_{(t)} = [(1 - R^2_1)/(N-2)] = [(1 - t^2)/(N-2)]$. For Model 2, the $R^2_2 = r^2_{ZX} = a^2$ and the squared standard error for the a coefficient is: $SE^2_{(a)} = [(1 - R^2_2)/(N-2)] = [(1 - a^2)/(N-2)]$. For both of these single-predictor models, there is no variance inflation due to collinearity.

Using the variance of a linear composite, the R^2 for a multiple (K) predictor standardized regression model can be calculated as: $R^2 = \mathbf{B}'\mathbf{R}_{XX}\mathbf{B}$: where $\mathbf{B}$ is a $K \times 1$ vector of standardized sample regression coefficients and $\mathbf{R}_{XX}$ is the $K \times K$ matrix of correlations among the predictors. For the $K = 2$ predictor Model 3, $\mathbf{B}' = [b \ t']$. Model 2 defines the correlation (i.e., collinearity) between the two predictors in Model 3 (X and Z). The 2×2 matrix of sample predictor correlations $\mathbf{R}_{XX}$ has ones on the diagonal and a as the off-diagonal elements. Thus, R^2 for Model 3 can be solved as:

$$R^2_3 = \mathbf{B}'\mathbf{R}_{XX}\mathbf{B} = \begin{bmatrix} b & t' \end{bmatrix} \begin{bmatrix} 1 & a \\ a & 1 \end{bmatrix} \begin{bmatrix} b \\ t' \end{bmatrix} = t'^2 + b^2 + 2abt' \quad (4)$$

The standard errors for each coefficient will be affected by their tolerance, which is related to a (i.e., the collinearity between the predictors X and Z). Specifically, tolerance is equal to $[1 - R^2_2] = [1 - a^2]$. The inflation of the variance for the b coefficient, Variance Inflation Factor (VIF), is the reciprocal of tolerance: $VIF = 1/[1 - a^2]$. With a two predictor model both predictors have the same tolerance and the squared standard errors for both the b and t' standardized coefficients are:

$$SE^2_{(t')} = SE^2_{(b)} = (1 - R^2_3) / [(N - 3)(1 - R^2_2)] = (1 - t'^2 + b^2 + 2abt') / [(N - 3)(1 - a^2)] \quad (5)$$

Thus, increases in the X - Z relationship (a) will reduce the tolerance of Z and inflate the variance (i.e., squared standard error) of the b coefficient in Model 3 by a factor of $VIF = 1/[1 - a^2]$.

The Sobel normal-theory product test is formed as:

$$z_s = \frac{ab}{\sqrt{b^2 SE^2_{(a)} + a^2 SE^2_{(b)}}} = \frac{ab}{\sqrt{\frac{b^2(1-a^2)}{(N-2)} + \frac{a^2(1-t'^2+b^2+2abt')}{(N-3)(1-a^2)}}} \quad (6)$$

With a in the numerator and a^2 and $(1 - a^2)$ in the denominator, if b and t' are held constant then the Sobel product test (z_s) is a nonlinear function of the a coefficient. The magnitude of z_s will increase with increases in a in certain ranges. In other ranges of the a coefficient, however, the magnitude of z_s will decrease with increases in a . Therefore, the statistical power of the product test will not necessarily increase with a .

Variance of the Sobel Product Test as a Function of the a and b Coefficients

The Sobel normal theory test (6) is employed to demonstrate trends in variance inflation as function of the a and b coefficients. The Expected Variance of the Sobel test of ab ($SE^2_{(ab)S}$) and the components that comprise it (i.e., $SE^2_{(a)}$; $SE^2_{(b)}$; $b^2SE^2_{(a)}$; $a^2SE^2_{(b)}$) were calculated for complete mediation models. Based on considerations discussed later, a sample size of $N = 100$ was used. With this fairly large fixed sample size, N operates basically as a constant in the components that make up $SE^2_{(ab)S}$ (see eq. 6), and thus, the trends reported would occur for any moderate-to-large sample size ($N \geq 40$). These variances were plotted as a function of b varied from 0 to 0.99 in increments of 0.01 for values of $a = 0.20, 0.50$, and 0.70 (right panels of Figure 2). Also, variances based on values of $b = 0.20, 0.50, 0.70$ were plotted against a varied from 0 to 0.99 (left panels of Figure 2); however, for clarity only $SE^2_{(ab)S}$, $SE^2_{(a)}$, and $SE^2_{(b)}$ were graphed. Across conditions, the products (ab) will be the same regardless of which component was varied. For example, whether a is fixed at 0.50 or b is fixed at 0.50, the product of a (or b) with the other 100 b (or a) = 0 to 0.99 values will result in the same ab product so that whether the graph is plotted with a (left panels of Figure 2) or with b (right panels of Figure 2) on the X-axis the trends are comparable. As can be seen in the right panels of Figure 2, the variance of the a coefficient ($SE^2_{(a)}$) is not affected by the magnitude of b . Although it is not graphed, it is noted that the $b^2SE^2_{(a)}$ component necessarily increases with b . The variance of the b coefficient ($SE^2_{(b)}$) and the $a^2SE^2_{(b)}$ component (not shown) decrease with increases in b . The net result is that error term of the Sobel normal-theory test ($SE^2_{(ab)S}$) increases with b to some degree for smaller values of $a = 0.2$ and 0.5 . For $a = 0.7$, $SE^2_{(ab)S}$ reduces slightly with increases in b . By contrast in the left panels of Figure 2, $SE^2_{(a)}$ and the $b^2SE^2_{(a)}$ component (not shown) decreases with increases in a ; however, the variance of b ($SE^2_{(b)}$), the $a^2SE^2_{(b)}$ component (not shown), and importantly, the standard error of the product ($SE^2_{(ab)S}$) increase dramatically with increases in a , especially as a approaches and exceeds b in magnitude. Thus, when the a coefficient becomes large, $SE^2_{(a)}$ becomes small and the magnitude of $SE^2_{(ab)}$ is dominated by the dramatic increase in $SE^2_{(b)}$ due to high collinearity (i.e., large values of a). Although the a coefficient is the effect of the independent variable (X) on the mediator (Z) in Model 2, by definition it is also the collinearity between the Z and X in Model 3. As a result, the expected variance of the product increases and the power of the Sobel normal-theory product test can be drastically affected by the magnitude of the a coefficient.

Predicted Power of the Sobel Normal-Theory Product Test

The Sobel normal theory test (6) is employed to demonstrate trends in statistical power for mediation analysis as function of the a and b coefficients. Although the statistical power for the Aroian and Goodman product tests and the difference test are not demonstrated, under many circumstances the statistical properties of these tests are quite similar, especially for larger sample sizes (MacKinnon et al., 1995). Thus, the power functions for all of these normal-theory tests of mediation will be affected by the magnitude of a (Beasley, 2012). A sample size of approximately $N \approx 100$ has been suggested in order to produce accurate estimates in simple mediation models (Hoyle and Kenny, 1999; MacKinnon, Lockwood, & Williams, 2004). Furthermore, for a significance level of $\alpha = 0.05$, a sample size of $N = 100$ provided a wide range power values in order to illustrate how power can decline with the a

coefficient. Smaller sample sizes yielded extremely flat power functions. Much larger sample sizes yielded power functions that reached the maximum of 1.0 for a large portion of a and b combinations.

Statistical power is a function of the absolute value of the product, $|ab|$, and the power functions are identical and symmetric regardless of the signs of a or b . Thus, the Sobel test statistic, z_S (6), was computed only for positive values of a and b with $N = 100$. Using a and b as sample realizations of their respective population parameters, a and β , z_S (6) serves as a non-centrality parameter, and thus, comparing a normal distribution centered at z_S to the central normal distribution critical value of $Z_{0.975} = 1.96$ yields the predicted statistical power for the Sobel test at a two-tailed significance level of $\alpha = 0.05$ under the assumption that ab is normally distributed. Given that the distribution of ab is known to deviate from normality, these predicted power values will be different compared to empirical power estimated from a simulation study, even if normally distributed error terms are generated for Models 2 and 3.

For complete mediation models ($H_{0(4)}$: $\tau' = 0$ is true), standardized values of both a and b can be large in absolute value ($|a|$ and $|b| > 0.99$); with any pattern of positive and negative values; however, if $a = \pm 1$ then b is undefined because X and Z would be completely collinear as predictors in Model 3 and the $\mathbf{R}_{XX}$ matrix would be singular. Figure 3 shows that with $N = 100$ predicted statistical power for the Sobel product test in a complete mediation model reaches a maximum, plateaus, and then eventually declines as a function of a . For smaller values of $b = 0.15 - 0.20$, predicted power reaches a maximum around $a \approx 0.30$ and gradually declines shortly thereafter. For moderate values of $b = 0.35 - 0.40$, power reaches a maximum around $a \approx 0.50$, but does not decline rapidly until approximately $a \approx 0.80$. For larger values of b , power tends to reach the maximum of approximately 1.00 and not decline until a is extremely large. For example, $b = 0.60$, power reaches 0.99 at $a = 0.50$ and does not go below 0.50 power until $a > 0.95$ (see Figure 5).

Figures 4 and 5 display these predicted power results as a function of the b coefficient and show that the power of the Sobel test in a complete mediation model always increases when b increases in magnitude. This increase in power continues throughout the range of a ; however, the power functions change for larger values of a . In Figure 4, predicted power for $a = 0.50$ plateaus at approximately $b \approx 0.55$. Figure 5 shows that for larger values of a , the point at which the Sobel test reaches maximum power of approximately 1.00 shifts out to larger values of b . To elaborate, for $a = 0.80$ the point at which power plateaus shifts out to approximately $b \approx 0.6$. And for very large values of $a = 0.90, 0.95$, and 0.98 the points at power which plateaus at 1.00 shifts out to approximately $b \approx 0.7, 0.8$, and 0.9 , respectively. Also, when $a = 0.6$ and $b = 0.8$ or when $a = 0.8$ and $b = 0.6$, the product $ab = 0.48$ and predicted power has basically reached its maximum value of 1.00 (Figure 5); however, when $a = 0.98$ and $b = 0.5$, the product is slightly larger ($ab = 0.49$), but the power is drastically lower at 0.20. Yet when $a = 0.30$ and $b = 0.5$, the power is 0.77, even though the product is considerably smaller $ab = 0.15$ than the previous scenario (see Figures 3 and 4). These findings are consistent with Hoyle and Kenny's (1999) conclusion that the power of the product test of is typically maximized when " b is somewhat larger than a ."

More statistical power with b being larger than a for a fixed ab product is also evident in a partial mediation model (see Figure 6). For partial mediation models, the value of t' places constraints on the values of a and b . The values of a , b , and t' must yield bivariate correlations among X , Y , and Z so that the 3×3 correlation matrix is positive-definite. Therefore, a value of $t' = 0.30$ was used to demonstrate the predicted power of the Sobel test as a function of X - Z collinearity (a coefficient) in a partial mediation model. Based on the power results in for complete mediation models and to meet the positive-definite criterion, z_s (6) was computed for positive values of a ranging from 0 to 0.99 and b ranging from 0 to 0.50. Figure 6 shows the power functions of the Sobel z -test (6) in this partial mediation model are similar to those in Figure 3 in that power can decline with increases in a . It also demonstrates that the Sobel product test can be slightly more powerful as tests of partial mediation. That is, compared to complete mediation ($t' \approx 0$), a partial mediation model ($t' = 0$) will explain more variation (eq. 4) and reduce the standard errors for the coefficients in Model 3 (eq. 5) and the product (eq. 6).

For both complete (Figure 3) and partial (Figure 6) mediation models, predicted power increases monotonically as a function of b (see Figures 4 and 5). Increases in the absolute value of a also inflate the standard error of the product tests (see Figure 2) due to the concomitant increase in collinearity. The variance inflation in the denominator of the Sobel test due to increases in the a coefficient eventually overtakes the increase of the effect size (ab) in the numerator, which results in a loss of statistical power. Therefore, the predicted power of the product test will decline at some value of a depending on the value of b , the sample size (N), and the significance level (α). These findings are consistent with the Fritz et al. (2012) conclusion that the magnitude of b is an important factor in power “stagnation” in that for larger values of b the point at which power begins to decline shifts to larger values of a ; however, power never reduces as a direct function of the b coefficient. By contrast, Figure 3 shows that for a given value of b there is a value of the a coefficient at which statistical power for detecting mediation will decline. Although these specific trends only apply to a sample size of $N = 100$ and a significance level of $\alpha = 0.05$, the power decline does occur for larger sample sizes with more stringent significance levels.

In this study, a moderate sample size of $N = 100$ was used throughout. In other simulation studies, much larger sample sizes consistent with mediation analyses of small effect sizes have been examined. For example, Fritz et al. (2012) investigated statistical power in mediation analysis with a sample size as large as $N = 2,500$. According to the tables in Fritz and MacKinnon (2007) a sample size of approximately $N \approx 500$ is sufficient to detect an effect size of approximately $ab \approx 0.02$ with 80% power at $\alpha = 0.05$ significance level. Using the methods previously described, predicted power as a function of the a coefficient for a complete mediation model was examined for larger sample sizes of $N = 500$ and 2,500 at significance levels of $\alpha = 0.05$, 0.01 and 0.001. Although not reported, the power decline did occur with these larger sample sizes, but depending on the significance level, the reduction in power was not noticeable until $b < 0.25$. Consistent with Fritz et al. (2012) the most dramatic reductions in power occurred when was the Z - Y relationship was very small ($b < 0.10$) for a significance level of $\alpha = 0.05$.

SIMULATION STUDY

Nonparametric bootstrap approaches applied to mediation analysis do not estimate the standard error of ab , and therefore, one might presume that they are not susceptible to this power decline phenomenon; however, the results reported in Fritz et al. (2012) as well as other studies (e.g., MacKinnon & Fritz, 2007) imply that bootstrap methods have similar power functions as other tests of mediation. To perform bootstrapping, there must be data, and thus, the power of bootstrap approaches cannot be derived from hypothetical parameter values as in the previous demonstrations. Furthermore, as compared to results from randomly sampled or randomly generated data where the ab distribution may deviate from normality, the predicted power rates depicted in Figures 3 through 6 may be inaccurate to some extent. Therefore, a simulation study was conducted to compare the empirical power of the Sobel test with the bootstrap approaches and to demonstrate that bootstrap methods applied to mediation analysis are not immune to the decline in statistical power with increases in the magnitude of the a coefficient. The NI method was not included because previous simulation studies (Fritz & MacKinnon, 2007; Fritz et al., 2012) have shown that the power functions of NI were very similar to PB under the conditions used in this study.

To build upon the previous results, a fixed sample size of $N = 100$ and a significance level of $\alpha = 0.05$ was used. To emulate a randomized experiment, the independent variable (X) was generated as a fixed binary variable (i.e., Treatment vs. Control). A balanced design with 50 cases per category was used to avoid Type 1 error inflation associated with heteroscedasticity (Zimmerman, 2013) and to maximize power assuming equal costs for the two conditions (Cochran, 1963, p. 145). The X variable was then standardized. The equation for Model 2 was used to generate data for Z . The errors (ε_2) were sampled from a unit normal distribution using the RANNOR function in SAS/IML. To set the expected variance of Z equal to one, the variance of the error term was set at $\text{Var}(\varepsilon_2) = (1 - \alpha^2)$, where α is the population parameter for the sample coefficient a . The equation for Model 3 was used to generate data for Y . The errors (ε_3) were generated as unit normal and to maintain the expected variance of one for Y , the error term variance was set at $\text{Var}(\varepsilon_3) = (1 - [\tau^2 + \beta^2 + 2\alpha\beta\tau])$, where β and τ are population regression parameters for the partial relationships of Z and X with Y (i.e., sample estimates b and t'), respectively.

This data generation process was replicated $R = 5,000$ times. For each replicate, regression analyses using Models 2 and 3 with intercepts to correct for random variations from the expected values of zero for Z and Y were performed. From each of the 5,000 sets of results, parameters estimates, a and b , and their standard errors were obtained. The product ab served as the estimate of the mediation effect and the Sobel test (6) and its associated p -value were computed. The null hypothesis of no mediation [$H_0: a\beta = 0$] was rejected if the Sobel test p -value < 0.05 . Also, for each of the $R = 5,000$ replicates, $B = 5,000$ bootstrap samples were randomly selected. Regression analyses using Models 2 and 3 were performed to obtain bootstrap estimates of the a and b coefficients, respectively. Then their product ab was computed. The mean of the 5,000 bootstrap distributions of ab served as the point estimate of ab . PB, BC, and BCA bootstrap estimates of the 95% CIs for ab were obtained. If the bootstrap CI did not contain zero then the null hypothesis was rejected.

Figure 7 shows empirical rejection rates (Type 1 Error and Power) of the Sobel normal-theory test (6), the percentile bootstrap (PB), and the BC bootstrap for testing mediation effects as a function of the a coefficient under a complete mediation model for a fixed population regression parameter of $\beta = 0.35$. Rejection rates for the BCA bootstrap were extremely similar to those of BC, and therefore, BCA results are not displayed for graphic clarity. Consistent with many other simulation studies, when the a coefficient was simulated to be greater than zero the PB, BC, BCA had more empirical power than the Sobel normal-theory test. When the a coefficient was simulated to be equal to zero, the null hypothesis of no mediation was true ($H_0: a\beta = 0$), regardless of the value of b . Under this null, the Sobel normal-theory test had a Type 1 Error rate of 2.2%, considerably below the 5% nominal significance level, which explains the generally lower power of the Sobel test compared to other methods. The PB had Type 1 Error rates (4.84%) very near the number of rejections expected for the $\alpha = 0.05$ significance level. Consistent with other studies (Cheung, 2007; Fritz et al., 2012; MacKinnon et al., 2004) both BC and BCA (not shown) had inflated Type 1 Errors with rejection rates of approximately 7.5%. Thus, even when normal theory assumptions of Models 2 and 3 hold, BC and BCA have inflated false positive rates (Fritz & MacKinnon, 2007; Fritz et al., 2012) and the increased probability of making a Type 1 error may not be worth the slight increase in statistical power. Therefore, PB would seem to be the preferable method in terms of balancing valid Type 1 error rates and statistical power.

It is important to note that Figure 7 demonstrates that the power functions of the nonparametric bootstrap methods for testing mediation are similar to those of the Sobel normal-theory test in that power eventually declines with increases in a . Specifically with b simulated to be 0.35 and $N = 100$, the empirical power started to decline at approximately $a \approx 0.40$ for all methods. Thus, under these conditions, power was maximized at an effect size of approximately $ab \approx 0.14$ and declined at larger effect sizes. Although consistent with many other simulation studies, taken collectively, the findings in this study clearly depict the decrease in power as a function of increases in a (see Figures 3, 6, and 7). These findings also imply that the power decline can be expected due to an increase in mediator collinearity and a resultant increase in the standard error of normal-theory product tests (Figure 2). Furthermore, this phenomenon was demonstrated to occur with nonparametric bootstrap approaches that do not directly estimate standard errors (Figure 7).

Based on these findings, the statistical properties of the bootstrap distributions were investigated. The Mean, Variance, Skew, and Kurtosis of a , b , and ab were calculated for each of the $R = 5,000$ simulated bootstrap distributions. These descriptive statistics were then averaged over the $R = 5,000$ replications. Table 1 reports these values for the ab product. The Skew and Kurtosis indicate that the bootstrap distributions tend to be slightly non-normal for these values of a and b ; however, there did not seem to be extreme departures from normality. To compare the values in Table 1 to previous results, Figure 7 shows that when simulating $b = 0.35$ and $a = 0.50$ ($ab = 0.175$) with $N = 100$, statistical power at $\alpha = 0.05$ for the Percentile Bootstrap in a complete mediation model was approximately 0.88 and power has started to decline with increases in a . The mean bootstrap estimate of ab (i.e., average of each bootstrap distribution of ab) across the 5,000 replications was 0.1757, near the expected value of 0.175. The standard deviation of the

5,000 bootstrapped ab means was 0.06261, near the expected $SE_{(ab)}$ of 0.06287. The average variance of the bootstrap samples was 0.00422, also near the variance of 0.00395 expected from the Sobel normal-theory test. As can be seen in Figure 7, when b is fixed at 0.35 and a was increased to 0.80 ($ab = 0.28$), power for Percentile Bootstrap declined to approximately 0.60, even though the magnitude of ab increased. The mean bootstrap estimate of ab across the 5,000 replications (0.28160) was again close to its expected value. The average variance was larger (0.01684) and close to the expected variance of 0.01653. For comparison, if the values of a and b are switched ($a = 0.35$; $b = 0.80$; $ab = 0.28$), the predicted power of the Sobel normal-theory test was near 0.90 (see Figures 3 and 4). Under this condition, the empirical power for the Percentile Bootstrap was slightly higher near 0.97. The mean bootstrap estimate of ab across the 5,000 replications was 0.28058, but the average variance was smaller (0.00627). Both values were close to their expected values. In another scenario where the product was fixed at $ab = 0.28$, but the coefficients were less disparate with b slightly larger than a ($a = 0.50$; $b = 0.56$), PB had virtually maximum power in this situation, the mean bootstrap estimate of ab showed little bias, and the average variance was smaller at 0.00494 and close to the expected variance of 0.00476. Overall, the results in Table 1 indicate that the variance of the bootstrap distributions closely approximates the variance of ab expected from the Sobel normal-theory test with a sample size of $N = 100$.

The distributional characteristics of the bootstrap samples shown in Table 1 indicate that the decline in power for bootstrap tests with increases in a are not attributable to bias in estimating the product or extreme non-normality in the bootstrap distributions. In comparing the first condition ($a = 0.50$; $b = 0.35$) to the fourth ($a = 0.50$; $b = 0.56$), increasing b did not appear to greatly affect the variance of the bootstrap samples. By contrast, comparing the second condition ($a = 0.80$; $b = 0.35$) to other conditions, the variance of the bootstrap samples may be 2 to 3 times larger when the a coefficient is larger than b , even though the product is the same.

Although further investigations are needed to make a more comprehensive conclusion, this implies that the dramatic increase in the variance of ab as a function of a demonstrated in Figure 2 also applies to using nonparametric bootstrap methods in mediation analysis. In practice, these inflated variances imply that the distribution of bootstrap estimates of ab have a wider range, and therefore, the upper and lower percentiles are more extreme and more likely to contain zero, thus leading to more non-rejections and less power. Although the nonparametric bootstrap approach may yield higher power than the normal-theory product tests, it is not a solution to the decline in power experienced when the a coefficient is substantially larger than the b coefficient. Also, given the similarity between the NI and PB power functions from other studies, it is reasonable to conclude that NI would also suffer from a decline in statistical power under these circumstances.

COMPUTATIONAL EXAMPLE

Stigler et al. (2006) reported the results of a mediation analysis for a controlled, group-randomized trial of an AIDS education program that was implemented with 814 school children in Tanzania. One of the main outcomes of interest was *Changes in Attitudes*

Towards People with AIDS. The authors presented eight single mediator models and found that four of the single mediator analyses indicated significant mediation. From their results, it can be derived that the effect of the intervention on the *Subjective Norms about Sexual Activity* mediator (i.e., a coefficient) was smaller than this mediator's association with the outcome (i.e., b coefficient). The effect of the intervention on the *Communication about HIV/AIDS* mediator was stronger than this mediator's association with the outcome ($a > b$).

To present a concrete example of power decline in mediation analysis, suppose this health education program evaluation study was replicated with $N = 100$ children randomized to Control and Intervention conditions (see Figure 1). Table 2 reports the Means, Standard Deviations (SD), correlation matrix, and the results of two separate regression-based mediation analyses: one with *Subjective Norms about Sexual Activity* (Z_1) and a second with *Communication about HIV/AIDS* (Z_2). The Appendix shows SAS code for generating and analyzing these data. These data were also submitted to the INDIRECT SAS macro (Preacher & Hayes, 2008) to obtain the nonparametric bootstrap CIs reported in Table 3.

In the unadjusted Model 1, the AIDS education program (X) had a statistically significant effect on the outcome, *Changes in Attitudes Towards People with AIDS* ($t = 0.1989$; $p = 0.0473$). In the first mediation analysis, the results from Model 2 show that the AIDS education program (X) had a statistically significant but relatively weak effect on the mediator, *Communication about HIV/AIDS* ($a_1 = 0.2734$, $p = 0.0059$). Model 3 from the first mediation analysis indicates that this mediator had a significant and relatively strong influence on the outcome in Model 3 ($b_1 = 0.4187$, $p < 0.0001$) and that the effect of the AIDS Education Program was reduced and no longer statistically significant ($t'_1 = 0.0844$, $p = 0.3731$). This resulted in a statistically significant product ($ab_1 = 0.1145$) using the Sobel normal-theory test ($p = 0.0167$) as well as the nonparametric bootstrap CIs (see Table 3). Thus, the effect of the AIDS Education Program on the outcome was mediated by *Communication about HIV/AIDS* (Z_1). In this case, the strong link between with the *Communication about HIV/AIDS* (Z_1) mediator and the *Changes in Attitudes Towards People with AIDS* outcome (i.e., Conceptual Theory) compensated for the relatively weak effect of the AIDS education program on the mediator (i.e., Action Theory).

In the first mediation analysis, the b_1 coefficient was larger than the a_1 coefficient and the statistical power to detect mediation was adequate even though the indirect (mediated) effect was actually smaller than the product in the second mediation analysis. In the second mediation analysis, the AIDS Education Program (X) had a statistically significant effect on the *Communication about HIV/AIDS* mediator ($a_2 = 0.6706$, $p < 0.0001$). Although the effect of the AIDS program on the outcome was reduced and non-significant ($t'_2 = 0.0502$, $p = 0.7050$) in mediation Model 3, the mediator's partial association with the outcome ($b_2 = 0.2217$, $p = 0.0968$) was also not significant. The product was not statistically significant by the Sobel test ($ab_2 = 0.1487$; $p = 0.0960$) or the PB method. As compared to PB, the BC and BCA CIs in the second mediation analysis did not contain zero indicating statistically significant mediation (see Table 3); however, these results may be a reflection of the inflated Type 1 error rates of these two procedures reported in this study as well as other investigations (e.g., Cheung, 2007, 2009; Fritz et al., 2012). In conclusion, the second mediation analysis does not provide conclusive evidence that the significant effects of the

AIDS education program on both the second mediator, *Subjective Norms about Sexual Activity* ($r_{XZ_2} = 0.2554$, $p = 0.0103$), and the outcome, *Changes in Attitudes Towards People with AIDS* ($t = 0.1989$; $p = 0.0473$), occur as a mediational process. This illustrates that even with strong Action Theory (i.e., effect of the AIDS education program on the *Subjective Norms about Sexual Activity* mediator), it is difficult to detect mediation without strong Conceptual Theory linking the mediator to the outcome.

Statistically, the AIDS education program was more effective in changing the second mediator (Z_2) and had a larger indirect (mediated) effect on the outcome ($ab_2 = 0.1487$; $p = 0.0960$) as compared to the first mediation analysis ($ab_1 = 0.1145$; $p = 0.0167$).

Paradoxically, only the smaller product in the first mediation analysis was statistically significant by all methods, thus illustrating the reduction in power when a is substantially larger than b . Table 1 shows that the correlation of the second mediator (Z_2) with the outcome (Y) was significant ($r_{YZ_2} = 0.2554$, $p = 0.0103$); however, in the mediation Model 3 that regressed Y on to both X and Z_2 , this mediator's partial association with the outcome was not statistically significant ($b_2 = 0.2217$, $p = 0.0968$). This lack of significance can be attributed to the relatively high collinearity of Z_2 with the X variable ($a_2 = 0.6706$), which results in a tolerance of 0.5503 and a VIF of 1.8172. This variance inflation can be seen in both the Sobel normal theory $SE_{(ab)S}$ (Table 2) and Bootstrap SE_{BS} (Table 3); both were larger in the second mediation analysis as compared to the first mediation analysis. Furthermore, the nonparametric bootstrap CIs in Table 3 were wider in the second mediation analysis, demonstrating that large values of the a coefficient can lead to increased variance and a wider range of values in bootstrap distributions of the ab product.

DISCUSSION

Although it would seem that the statistical power for testing the ab product in mediation analysis would increase with the magnitude of the X - Z relationship (a coefficient), power can actually be reduced when a increases in absolute value. In the conditions presented, power typically reached a maximum, plateaued, and then declined with increases in $|a|$. Power did not decline with increases in $|b|$ (see Figures 4 and 5), but the value of a at which the power begins to decline does depend on the magnitude of b , the sample size (N), and significance level (α). For mediation models with stronger Z - Y partial relationships (b), the point at which power reaches a maximum was shifted to a larger value of $|a|$. This is consistent with the conclusion that b is an important factor in the power decline (Fritz et al., 2012); however, this study shows that although influenced by the magnitude of b , the increase the variance of the ab product and the resultant reduction in statistical power for both normal-theory and nonparametric bootstrapping approaches to testing mediation are strongly associated with the magnitude of the a coefficient.

In this study, a moderate sample size of $N = 100$ and significance level of $\alpha = 0.05$ was used throughout to show how power functions can decline with the a coefficient. For larger sample sizes ($N = 500$) and standard significance levels (e.g., $\alpha = 0.05$ or 0.01), this power decline with increases in the a coefficient may be less of a concern unless very small b coefficients are expected (Fritz et al., 2012). Although more comprehensive studies are needed, if mediation is tested at more stringent significance levels (e.g., $\alpha = 0.001$), then

power decline with increases in the a coefficient may be of concern even with larger b coefficients.

As discussed in the next section, the decline in statistical power should be expected because by definition the a coefficient in Model 2 is the collinearity between the independent variable (X) and the mediator (Z) in Model 3. The decrease in power can be attributed to increases in the standard error of normal-theory product tests due to collinearity. Yet, this phenomenon also occurred with nonparametric bootstrapping approaches that do not directly estimate standard errors because the variance of the bootstrap distribution of ab closely approximates the variance expected from the Sobel normal-theory test (see Table 1). In conjunction, these findings demonstrate that a decline in statistical power for mediation analysis is not an unexpected anomaly, but rather a phenomenon that can be derived from the formula for the Sobel normal-theory test (6). Next, implications for statistical analysis, power and sample size estimation, and the design of applied research examining mediational processes are discussed.

Implications for Statistical Analysis

The variance of the ab product was demonstrated to dramatically increase with increases in the a coefficient, especially when $a > b$. This is because in order to estimate the partial relationship of Z with Y conditional on X (b coefficient), both X and Z are included in Model 3. Thus, by definition the collinearity between Z and X in Model 3 increases with a in Model 2. The resultant variance inflation for the b coefficient in Model 3 and increase in the standard error ($SE_{(ab)S}$) of Sobel product test (6) is a function of a coefficient in Model 2. In turn, the variance inflation in the denominator of the Sobel test (6) due increases in collinearity (i.e., a coefficient in Model 2) eventually outweighs the increase of the effect size (ab) in the numerator and results in a loss of statistical power. Theoretically, the VIF in Model 3 is an index that represents how much $SE^2_{(b)}$ is increased by the collinearity between Z and X compared to a situation where collinearity is zero (i.e., tolerance and VIF = 1). However, if collinearity in mediation Model 3 is zero, then $a = 0$ in Model 2, and by definition, there is no mediation effect. This presents a quandary in that researchers investigating mediational processes would not want collinearity in Model 3 (a coefficient) to be extremely large, but they also would not want the a coefficient in Model 2 to be zero.

Implications for Conducting Power and Sample Size Calculations

The finding that increasing the magnitude of the effect size (ab) by increasing $|a|$ can reduce statistical power has important implications for researchers conducting power and sample size calculations for mediation analyses. For most linear regression-based statistical procedures, power increases with effect size. For example, the variance of a mean difference is independent of the magnitude of its effect size and statistical power increases with the magnitude of the effect size. The standard error of a correlation coefficient reduces with the magnitude of the effect size, and thus, statistical power also increases with the magnitude of the effect size. This principle does not apply to linear regression-based mediation analysis because the variance of the ab product is not independent of its magnitude and is a complicated function of the associations of the mediator with the outcome (b coefficient) and the independent variable (a coefficient), the distribution of the ab product, as well as

sample size (N). Therefore, the standard error of the product may actually increase with increases in the effect size, which may result in reduced statistical power or increased sample size requirements. This suggests that if a causal agent is expected to have a large effect on the mediator (a coefficient is expected to be large), the statistical power for tests of mediation may be less than one would expect based on the magnitude of the ab product. Therefore, power and sample size calculations should be conducted systematically to account for the potential decline in power, rather than assuming power will generally increase with a larger effect size.

Implications for Applied Research

This paper demonstrates that the decline in power as the a coefficient and effect size (ab) increase is an explainable statistical paradox attributable to the variance inflation caused by the collinearity between the mediator (Z) and the independent variable (X). For researchers conducting non-randomized studies or studies that do not involve an intervention, but who are interested in causal mechanisms, this implies that a strong X - Z relationship (a coefficient) may interfere with detecting mediation and understanding underlying mediational processes. This issue of declining power is also important for the common situation where X is an intervention and researchers employ Conceptual and Action Theory to design programs to indirectly affect outcomes through mediational processes.

The results show that power always increased with b (Figures 4 and 5), stressing the importance of the Conceptual Theory underlying the association between potential mediators and the outcome. As b becomes greater in magnitude, the value of a at which power begins to decline becomes larger and possibly unlikely to occur in practice. Thus, in the presence of strong Conceptual Theory (larger b coefficients), the power decline becomes less of a concern. In the situation where investigators are designing novel interventions in the presence of well-established Conceptual Theory, there may be expectations about the magnitude of the b coefficient and power and sample size calculations should be conducted for a range of a coefficients in order to determine situations where statistical power may decline.

Researchers often employ Action Theory to design interventions that maximize treatment effects on potential mediators (a coefficient). Thus, if the designed intervention is successful and results in a large a coefficient, whereas the b coefficient is medium-to-small in magnitude, then the effect of the intervention on the intermediate variable can paradoxically reduce the power to detect mediational effects and decrease the ability to discover and understand mediational processes. In the presence of strong Action Theory, the effect of the intervention on the mediator (a coefficient) may be expected to be large, and depending on the sample size and significance level, the b coefficient must be of sufficient magnitude in order to yield adequate statistical power. Therefore, even when there is strong Action theory, researchers should still rely on well-established Conceptual Theory in order to measure mediators that are strongly related to the outcome (i.e., larger b values); otherwise, mediation analyses may have low statistical power or exorbitant sample sizes may be necessary to detect mediation. This creates a dilemma in that theoretically attempting to develop an intervention that has strong effects on the mediator through well-developed

Action Theory may be of little consequence or may even be counter-productive in terms of detecting mediation when there is not strong *Conceptual Theory*. Thus, when the outcome of interest is difficult to affect directly and can only be influenced indirectly through a mediational process, Conceptual Theory is paramount to researchers designing interventions to affect intermediate variables. Without strong Conceptual Theory linking the intermediate variable to the outcome, one cannot expect an intervention to affect an outcome through a mediational process, even when an intervention has a strong effect on the intermediate variable.

Acknowledgments

This work was partially supported by National Institutes of Health P60-AR048095.

REFERENCES

- Albert JM. Mediation analysis via potential outcomes models. *Statistics in Medicine*. 2008; 27:1282–1304. [PubMed: 17691077]
- Allison DB, Heo M, Schork NJ, Wong SL, Elston RC. Extreme selection strategies in gene mapping studies of oligogenic quantitative traits do not always increase power. *Human Heredity*. 1998; 48:97–107. [PubMed: 9526169]
- Aroian LA. The probability function of the product of two normally distributed variables. *Annals of Mathematical Statistics*. 1947; 18:265–271.
- Barger JL, Kayo T, Pugh TD, Prolla TA, Weindruch R. Short-term consumption of a resveratrol-containing nutraceutical mixture mimics gene expression of long-term caloric restriction in mouse heart. *Experimental Gerontology*. 2008; 43(9):859–866. [PubMed: 18657603]
- Baron RM, Kenny DA. The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*. 1986; 51:1173–1182. [PubMed: 3806354]
- Beasley TM. Power of Product Tests of Mediation as a Function of Mediator Collinearity. *Multiple Linear Regression Viewpoints*. 2012; 38(2):17–23.
- Biesanz JC, Falk CF, Savalei V. Assessing mediational models: Testing and interval estimation for indirect effects. *Multivariate Behavioral Research*. 2010; 45:661–701.
- Bollen KA, Stine RA. Direct and indirect effects: Classical and bootstrap estimates of variability. *Sociological Methodology*. 1990; 20:115–140. doi: 10.2307/271084.
- Chen, H-T. *Theory-Driven Evaluations*. Sage; Newbury Park, CA: 1990.
- Cheung MW-L. Comparison of approaches to constructing confidence intervals for mediating effects using structural equation models. *Structural Equation Modeling*. 2007; 14:227–246. doi: 10.1080/10705510701293627.
- Cheung MW-L. Comparison of methods for constructing confidence intervals of standardized indirect effects. *Behavior Research Methods*. 2009; 41:425–438. doi: 10.3758/BRM.41.2.425. [PubMed: 19363183]
- Cochran, WG. *Sampling Techniques*. Wiley; New York: 1963.
- Cox TL, Ard JD, Beasley TM, Fernández JR, Howard VJ, Affuso O. Body Image as a Mediator of the Relationship Between Body Mass Index and Weight-Related Quality of Life in Black Women. *Journal of Women's Health*. 2011; 20(10):1573–1578.
- Goodman LA. On the exact variance of products. *Journal of the American Statistical Association*. 1960; 55:708–713.
- Efron, B.; Tibshirani, RJ. *An introduction to the bootstrap*. Chapman & Hall/CRC.; Boca Raton, FL: 1993.
- Fairchild, A. j.; MacKinnon, D. p. *A General Model for Testing Mediation and Moderation Effects*. *Prevention Science*. 2009; 10(2):87–99. [PubMed: 19003535]

- Forestell, CA.; Mennella, JA. More than just a pretty face. The relationship between infant's temperament, food acceptance, and mothers' perceptions of their enjoyment of food. *Appetite*. 2012. <http://dx.doi.org/10.1016/j.appet.2012.03.005>
- Fritz MS, MacKinnon DP. Required sample size to detect the mediated effect. *Psychological Science*. 2007; 18:233–239. doi: 10.1111/j.1467-9280.2007.01882.x. [PubMed: 17444920]
- Fritz MS, Taylor AB, MacKinnon DP. Explanation of Two Anomalous Results in Statistical Mediation Analysis. *Multivariate Behavioral Research*. 2012; 47(1):61–87. <http://dx.doi.org/10.1080/00273171.2012.640596>. [PubMed: 24049213]
- Hanks LJ, Casazza K, Willig AL, Cardel MI, Beasley TM, Fernández JR. Associations among calcium intake, resting energy expenditure, and body fat in a multiethnic sample of children. *Journal of Pediatrics*. 2010; 157(3):473–478. [PubMed: 20400090]
- Hayes AF. Beyond Baron and Kenny: Statistical mediation analysis in the new millennium. *Communication Monographs*. 2009; 76:408–420.
- Hoyle, RH.; Kenny, DA. Statistical power and tests of mediation. In: Hoyle, RH., editor. *Statistical strategies for small sample research*. Sage; Newbury Park: 1999. p. 195–222.
- Lipsey, MW. Theory as method: Small theories of treatments. In: Sechrest, LB.; Scott, AG., editors. *Understanding Causes and Generalizing About Them*. Jossey-Bass; San Francisco: 1993. p. 5–38.
- Lomnicki ZA. On the distribution of the products of random variables. *Journal of the Royal Statistical Society Series B*. 1967; 29:513–523.
- MacCorquodale K, Meehl PE. On a distinction between hypothetical constructs and intervening variables. *Psychological Review*. 1948; 55:95–107. [PubMed: 18910284]
- Meyer-Lindenberg A, Weinberger DR. Intermediate phenotypes and genetic mechanisms of psychiatric disorders. *National Review of Neuroscience*. 2006; 7:818–827.
- Munafò MR. Candidate gene studies in the 21st century: meta-analysis, mediation, moderation. *Genes, Brain, and Behavior*. 2006; 5:3–8.
- MacKinnon, DP. *Introduction to Statistical Mediation Analysis*. Erlbaum; Mahwah NJ: 2008.
- MacKinnon DP, Fritz MS, Williams J, Lockwood CM. Distribution of the product confidence limits for the indirect effect: Program PRODCLIN. *Behavior Research Methods*. 2007; 39:384–389. [PubMed: 17958149]
- MacKinnon DP, Lockwood CM, Hoffman JM, West SG, Sheets V. A comparison of methods to test the significance of the mediated effect. *Psychological Methods*. 2002; 7(1):83–104. [PubMed: 11928892]
- MacKinnon DP, Lockwood CM, Williams J. Confidence limits for the indirect effect: Distribution of the product and resampling methods. *Multivariate Behavioral Research*. 2004; 39(1):99–128. [PubMed: 20157642]
- MacKinnon DP, Warsi G, Dwyer JH. A simulation study of mediated effect measures. *Multivariate Behavioral Research*. 1995; 30:41–62. [PubMed: 20157641]
- Mallinckrodt B, Abraham WT, Wei M, Russell DW. Advances in testing the statistical significance of mediation effects. *Journal of Counseling Psychology*. 2006; 53:372–378. doi: 10.1037/0022-0167.53.3.372.
- McGuigan, K.; Langholz, B. A note on testing mediation paths using ordinary least-squares regression. 1988. Unpublished note
- Meulders A, Vansteenwegen D, Vlaeyen JW. Women, but not men, report increasingly more pain during repeated (un)predictable painful electrocutaneous stimulation: evidence for mediation by fear of pain. *Pain*. 2012 <http://dx.doi.org/10.1016/j.pain.2012.02.005>.
- Pedhazur, EJ. *Multiple regression in behavioral research*. Harcourt Brace; Orlando, FL: 1997.
- Pituch KA, Stapleton LM, Kang JY. A comparison of single sample and bootstrap methods to assess mediation in cluster randomized trials. *Multivariate Behavioral Research*. 2006; 41:367–400. doi: 10.1207/s15327906mbr4103_5.
- Preacher KJ, Hayes AF. Asymptotic and resampling methods for estimating and comparing indirect effects. *Behavior Research Methods*. 2008; 40:879–891. [PubMed: 18697684]
- Popescu M, Boisjoly H, Schmaltz H, Kergoat MJ, Rousseau J, Moghadaszadeh S, Djafari F, Freeman EE. Explaining the Relationship Between Three Eye Diseases and Depressive Symptoms in Older

- Adults. *Investigative Ophthalmology & Visual Science*. 2012 [Epub ahead of print] doi: 10.1167/iops.11-9330.
- Shrout PE, Bolger N. Mediation in experimental and nonexperimental studies: New procedures and recommendations. *Psychological Methods*. 2002; 7:422–445. [PubMed: 12530702]
- Sobel, ME. Asymptotic confidence intervals for indirect effects in structural equation models. In: Leinhardt, S., editor. *Sociological Methodology* 1982. American Sociological Association; Washington DC: 1982. p. 290-312.
- Springer MD, Thompson WE. The distribution of products of independent random variables. *SIAM Journal of Applied Mathematics*. 1966; 14:511–526. doi: 10.1137/0114046.
- Stoltenberg SF, Twitchell GR, Hanna GL, Cook EH, Fitzgerald HE, Zucker RA, Little KY. Serotonin transporter promoter polymorphism, peripheral indexes of serotonin function, and personality measures in families with alcoholism. *American Journal of Medical Genetics*. 2002; 114:230–234. [PubMed: 11857587]
- Varnhagen CK, Zumbo BD. CAI as an adjunct to teaching Introductory Statistics: Affect mediates learning. *Journal of Educational Computing Research*. 1990; 6(1):29–40.
- Veitch J, van Stralen MM, Chinapaw MJ, Te Velde SJ, Crawford D, Salmon J, Timperio A. The neighborhood social environment and body mass index among youth: a mediation analysis. *International Journal of Behavior, Nutrition & Physical Activity*. 2012; 9(1):31.
- Wilkins LK, Girarda TA, Cheyne JA. Anomalous bodily-self experiences among recreational ketamine users. *Cognitive Neuropsychiatry*. 2012 DOI: 10.1080/13546805.2012.663162.
- Williams J, MacKinnon DP. Resampling and distribution of the product methods for testing indirect effects in complex models. *Structural Equation Modeling*. 2008; 15:23–51. doi: 10.1080/10705510701758166. [PubMed: 20179778]
- Zieffler, AS.; Harring, JR.; Long, JD. *Comparing groups: Randomization and bootstrap methods using R*. Wiley; Hoboken, NJ: 2011.
- Zimmerman DW. Heterogeneity of variance and biased hypothesis tests. *Journal of Applied Statistics*. 2013; 40(1):169–193.

Figure 1.
Path Analysis Model of Mediation.

Figure 2.

Variance of a ($SE^2_{(a)}$), b ($SE^2_{(b)}$), and ab ($SE^2_{(ab)}$) coefficients as a function of the a and b coefficients in a complete mediation model with $N = 100$.

Figure 3.

Power of Sobel z-test as a function of the X-Z relationship (a) for a Complete Mediation Model at two-tailed $\alpha = 0.05$ significance level with $N = 100$.

Figure 4.

Power of Sobel z -test for $a = 0.10$ to 0.50 as a function of the Y - Z partial relationship (b) for a Complete Mediation Model at two-tailed $\alpha = 0.05$ significance level with $N = 100$.

Figure 5.

Power of Sobel z -test for $a = 0.60$ to 0.98 as a function of the Y - Z partial relationship (b) for a Complete Mediation Model at two-tailed $\alpha = 0.05$ significance level with $N = 100$.

Figure 6.

Power of Sobel z -test as a function of the X - Z relationship (a) for a Partial Mediation Model at two-tailed $\alpha = 0.05$ significance level with $N = 100$.

Figure 7. Power of Sobel z -test, the Percentile Bootstrap, and the Bias-Corrected Bootstrap as a function of the X - Z relationship (a) for a fixed value of $b = 0.35$ in a Complete Mediation Model at two-tailed $\alpha = 0.05$ significance level with $N = 100$.

Table 1Descriptive Statistics for ab across $R = 5,000$ Replicated Bootstrap Samples ($B = 5,000$).

Expected Values				Bootstrap Descriptive Statistics				
a	b	ab	Variance of ab ($SE_{(ab)S}$) *	Average of Bootstrap ab (SD)	Average Variance of Bootstrap ab	Average Skew of Bootstrap ab	Average Kurtosis of Bootstrap ab	Empirical Power of PB
0.50	0.35	0.175	0.00395 (0.06287)	0.17573 (0.06261)	0.00422	0.34618	0.33042	0.8835
0.80	0.35	0.280	0.01653 (0.12858)	0.28160 (0.12694)	0.01684	0.06671	0.14638	0.6010
0.35	0.80	0.280	0.00625 (0.07905)	0.28058 (0.07857)	0.00627	0.12294	0.07424	0.9655
0.50	0.56	0.280	0.00476 (0.06898)	0.28075 (0.06851)	0.00494	0.33631	0.20049	0.9990

Note:

* The Expected Variance of ab and $SE_{(ab)S}$ are computed from the denominator of the Sobel test in equation 6.

Table 2

Descriptive Statistics for Computational Example.

Variable	Means (<i>SD</i>)		Correlations	
	Control (<i>n</i> = 50)	Intervention (<i>n</i> = 50)	<i>Z</i> ₁ (<i>p</i> -value)	<i>Z</i> ₂ (<i>p</i> -value)
<i>Y</i>	9.36 (3.02)	10.56 (2.96)	0.4420 (<i>p</i> < 0.0001)	0.2554 (0.0103)
<i>Z</i> ₁	33.64 (5.21)	36.38 (4.50)		0.3499 (0.0004)
<i>Z</i> ₂	9.56 (1.67)	12.36 (1.45)		
	<i>a</i>	<i>b</i>	<i>ab</i>	<i>t'</i>
Mediation Analysis 1 (<i>SE</i>)	0.2734 (0.0972) <i>p</i> = 0.0059	0.4189 (0.0943) <i>p</i> < 0.0001	0.1145 (0.0479) <i>p</i> = 0.0167	0.0844 (0.0943) <i>p</i> = 0.3731
Mediation Analysis 2 (<i>SE</i>)	0.6706 (0.0749) <i>p</i> < 0.0001	0.2217 (0.1322) <i>p</i> = 0.0968	0.1487 (0.0893) <i>p</i> = 0.0960	0.0502 (0.1322) <i>p</i> = 0.7050

Note: *Y* = "Changes in Attitudes Towards People with AIDS." *X* = AIDS Education Program *Z*₁ = Subjective Norms about Sexual Activity." *Z*₂ = "Communication about HIV/AIDS." By definition, $t = r_{YX} = 0.1989$ ($p = 0.0473$) for both Models. For Mediation Analysis 1, $a_1 = r_{XZ1} = 0.2734$. For Mediation Analysis 2, $a_2 = r_{XZ2} = 0.6706$. Therefore, these values are not reported in the Correlation Matrix.

Table 3

Normal-Theory and Nonparametric Bootstrap Results for Computational Example.

	Mediation Analysis 1			Mediation Analysis 2		
	<i>ab</i>	Lower	Upper	<i>ab</i>	Lower	Upper
Sobel	0.1145	0.0206	0.2084	0.1487	-0.0263	0.3237
Bootstrap	0.1136	$SE_{BS} = 0.0450$		0.1466	$SE_{BS} = 0.0784$	
Percentile		0.0365	0.2117		-0.0025	0.3060
BC		0.0436	0.2245		0.0025	0.3119
BCA		0.0428	0.2236		0.0049	0.3145

Note: SE_{BS} = Standard Error of the Bootstrap Distribution. BC = Bias-Corrected. BCA = Accelerated and Bias Corrected. SE_{BS} = Standard Error of the Bootstrap Distribution. Because bootstrap estimation involves random resampling, the results can change each time the same analysis is performed. In order to reproduce the same results each time the bootstrap methods were used a fixed seed of 10133857 was set in the INDIRECT macro.