

Power of Product Tests of Mediation as a Function of Mediator Collinearity

T. Mark Beasley

University of Alabama at Birmingham

In a mediation model, the a coefficient represents the relationship of the independent variable (X) to the mediator (Z) and the b coefficient represents the partial relationship of the mediator (Z) to the outcome (Y). The amount of mediation can be estimated as the product of the two sample coefficients (ab). Although it would seem that an increase in a would increase statistical power for product tests, it also increases the standard error of the product tests due to the necessary increase in the collinearity between X and Z . Thus, power can actually decline with increases in a .

The first and perhaps the most popular approach for assessing mediation is to fit a sequence of linear regression equations (Baron & Kenny, 1986). Let Y_i denote the response for individual i , Z_i an intermediate (potential mediator) variable, and X_i a potential causal factor of interest for individual i . Since all regression based tests for mediation can be expressed in terms of standardized regression coefficients (Pedhazur, 1997; MacKinnon et al., 1995), the three separate linear models can be expressed as:

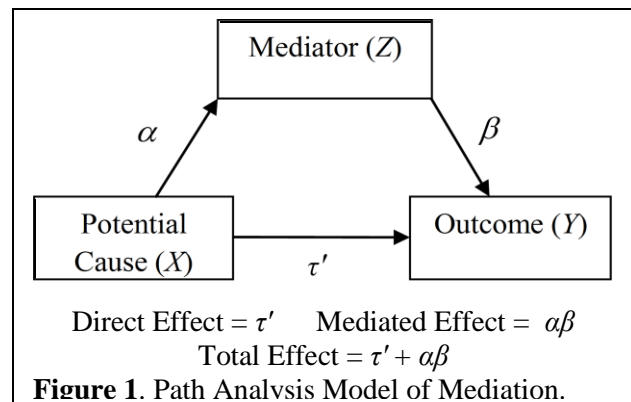
$$Y_i = \tau X_i + \varepsilon_{1i} \quad (1)$$

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

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

where τ , α , β , and τ' represent regression parameters estimated by sample coefficients t , a , b , and t' , respectively. The errors (ε 's) for each model are assumed to be normally distributed, uncorrelated, and to have mean zero and constant variance. Baron and Kenney (1986) consider the intermediate variable Z to be a mediator if: (1) $H_{0(1)}$: $\tau = 0$ is rejected (Y is associated with X); (2) $H_{0(2)}$: $\alpha = 0$ is rejected (Z is associated with X); and (3) $H_{0(3)}$: $\beta = 0$ is rejected (Y is associated with the Z conditional on X). The satisfaction of these conditions indicates that there is a mediating effect of X on Y through Z . If all three of these steps are met and $H_{0(4)}$: $\tau' = 0$ is rejected, then the data are consistent with the hypothesis that variable Z *partially* mediates the X - Y relationship, and if $H_{0(4)}$: $\tau' = 0$ is not rejected, then *complete mediation* is indicated.

Several authors have provided multiple examples of using a path analytic approach to solve the mediation analysis regression models simultaneously (e.g., Schumacker & Lomax, 1996; MacKinnon, 2008). In the path analytic approach, the relationship between an independent variable, X , and the outcome, Y , is decomposed into direct and indirect (mediated) effects as shown in Figure 1. The amount of mediation (*indirect effect*) is defined as the reduction of the effect of the initial variable on the outcome, which can be estimated in two ways: (1) the difference between two sample regression coefficients [$t - t'$] or (2) the product of two sample regression coefficients [ab]. The difference in parameters is theoretically the same as the product of the effect of X on Z times the effect of Z on Y ; thus it holds that $\alpha\beta \approx \tau - \tau'$. MacKinnon, Warsi, and Dwyer (1995) demonstrated that the two sample estimates of mediation are identical (i.e., $ab = [t - t']$) when (1) multiple regression (or path analysis) is used, (2) there are no missing data, and (3) the same covariates are in the equation.



Product Tests of Mediation

Currently dozens of methods for testing for mediation have been proposed with no consensus on which is best (Albert, 2008; MacKinnon et al., 2002). From Baron and Kenny (1986) demonstrating *complete mediation* requires rejecting the first three null hypotheses and not rejecting the fourth. Other methods assess mediation by performing a single test of whether the association between X and Y is *significantly smaller* when Z is controlled (i.e., $H_0: [\tau - \tau'] = 0$), but do not require the X - Y partial

relationship (t') to be non-significant (MacKinnon et al., 1995; McGuigan & Langholtz, 1988). Product tests use ab as the estimate of the mediation effect and perform a single test of $H_0: \alpha\beta = 0$. Sobel (1982) proposed a test that presumes that α and β are independent, and thus, the squared standard error (SE) for the Sobel test is a weighted composite the squared standard errors for a and b coefficients, $SE^2_{(ab)S} = b^2SE^2_{(a)} + a^2SE^2_{(b)}$. Other tests include the covariance of the a and b coefficients, which is typically small. For example, the Aroian (1947) test adds this covariance term, $SE^2_{(ab)A} = b^2SE^2_{(a)} + a^2SE^2_{(b)} + SE^2_{(a)}SE^2_{(b)}$, which was recommended by Baron and Kenny (1986) because it does not make the unnecessary assumption that the product of $SE^2_{(a)}$ and $SE^2_{(b)}$ is vanishingly small. The Goodman (1960) test subtracts this covariance term, $SE^2_{(ab)G} = b^2SE^2_{(a)} + a^2SE^2_{(b)} - SE^2_{(a)}SE^2_{(b)}$, to form an unbiased estimate of the variance of the mediated effect; however, this can yield negative variance estimates. All three of these approaches divide ab by the $SE_{(ab)}$ to form a z -statistic and use to the unit normal distribution to obtain p -values or calculate confidence intervals. The Sobel and Aroian tests have performed best in simulation studies (MacKinnon et al., 1995). Because most methods for testing mediation differ on how to define standard errors and confidence intervals, bootstrapped tests (Shrout & Bolger, 2002) and confidence intervals (Preacher & Hayes, 2008) are increasingly popular approaches; however, the Sobel normal theory test arguably remains one of the most widely-used test in mediation analysis.

Data Example

Following the framework of a cost-benefit analysis presented by Leech and Onwuegbuzie (2004), two indices were implemented. The first index related to the cost per level of effectiveness (CE), where C was the direct cost of the program and E was the practical effect measured in terms of the raw difference in testing points (i.e., ISAT mathematics and reading). Note: the practical effect of the effect size measure in terms of testing points gained was based on average standard deviations from sample data trends found for ISAT mathematics elementary = 28.04, mathematics MS = 27.83, reading elementary = 27.51, and reading MS = 24.27 (Consortium on Chicago School Research, 2007).

The Appendix presents data for a simple example of a single mediator model with $N = 20$ cases. Suppose X is a binary variable representing a randomized treatment (e.g., treatment vs control) that has been standardized to have a mean of zero and variance of one. That is, for a two-group balanced design, effect codes of $\pm \sqrt{(N - 1)/N}$ are assigned to indicate group membership. Also suppose that Z and Y are standardized continuous variables that represent the mediator and outcome, respectively. Since all regression based tests for mediation can be expressed in terms of standardized regression coefficients and bivariate correlations (Pedhazur, 1997; MacKinnon et al., 1994), the upper diagonal of Table 1 reports the correlation matrix among these three variables and displays the standardized mediation coefficients from Regression Models 2 and 3 with their SE s (in parentheses) on the lower diagonal. Note that the bivariate correlations of X with Y and Z in Table 1 are equal to the standardized coefficients for Models 1 (t) and 2 (a), respectively. The Appendix also contains SAS code and output for these analyses.

As can be seen the product of a and b is $ab = 0.58422 * 0.46755 = 0.273152$ and the estimated Total Effect [$t = (t' + ab) = (0.25778 + 0.273152) = 0.53093$] reproduces the bivariate X - Y correlation and demonstrates that $ab = [t - t']$. The Sobel squared standard error is $SE^2_{(ab)S} = (0.46755^2 * 0.19129^2) + (0.58422^2 * 0.22643^2) = 0.0254984$; thus the $SE_{(ab)S} = \sqrt{0.0254984} = 0.15968217$, and the $z_s = (0.273152 / 0.15968217) = 1.71059838$ with a two-tailed p -value = 0.08715528. The Aroian squared standard error adds the ab covariance term and is equal to $SE^2_{(ab)A} = 0.0254984 + (0.19129^2 * 0.22643^2) = 0.02737448$, which results in $z_A = 1.65094096$ with a two-tailed p -value = 0.09875063. The Goodman squared standard error subtracts this covariance term and is equal to $SE^2_{(ab)G} = 0.0254984 - (0.19129^2 * 0.22643^2) = 0.02362231$, which results in $z_G = 1.77722861$ with a two-tailed p -value = 0.07553063. All three test are not statistically significant at the $\alpha = 0.05$ level.

Table 1. Correlation Matrix (Upper Diagonal) and Standardized Mediation Coefficients (Lower Diagonal).

	Y	Z	X
Y	1.0000	0.61815	$t = 0.53093$
Z	$b = 0.46755$ (0.22643)	1.0000	$a = 0.58422$
X	$t' = 0.25778$ (0.22643)	$a = 0.58422$ (0.19129)	1.0000

Power of Product Tests

Power and sample size calculations for mediation analyses present a challenge because the power functions for the product tests depend on the strength of the associations of the mediator (Z) with the outcome (b), and the independent variable (a) and the distribution of their product. Hoyle and Kenney (1999, p. 201) briefly discuss power of testing $H_{0(3)}: \beta = 0$ and $H_{0(4)}: \tau' = 0$ as a function of the collinearity of X and Z . As they describe, the more variance in Z explained by X , the less variance in Z there is to contribute to uniquely explaining Y . That is, as a becomes large at some point b can be reduced. To elaborate, suppose an extreme example where $a = 1$ in a standardized Model 2. In this case, X explains 100% of Z and therefore Z cannot uniquely account for Y (i.e., b would have to be zero due to the complete collinearity). Hoyle and Kenney state that the degree a affects the power of product tests is less clear because both a and b contribute to the product; however, they do not discuss this issue in terms of the SE s of the product test statistics and the tolerance (i.e., collinearity) of the predictors in Model 3.

It would seem that the power of product tests would increase with a , because the numerator, ab , increases and part of the denominator, $SE_{(a)}^2 = (1 - a^2)$, decreases. What has not been noted, however, is that increases in a lead to increases in $SE_{(b)}^2$ in the denominator of the product tests due to the necessary increase in the collinearity between X and Z . Thus, power can actually be reduced by increasing a . To explicate this issue, again assume all variables are standardized, and thus, all regression based tests for mediation can be expressed in terms of standardized regression coefficients and bivariate correlations. For Model 1, the explained variation, $R^2_1 = t^2 = (t' + ab)^2$, the Mean Square Error (MSE) will be $MSE_1 = (1 - R^2_1)(N-1)/(N-2) = (1 - t^2)(N-1)/(N-2)$ and the squared standard error for the t coefficient is $SE_{(t)}^2 = [(1 - R^2_1)/(N-2)] = [(1 - t^2)/(N-2)]$. For Model 2, the $R^2_2 = a^2$ and the squared standard error for the a coefficient is $SE_{(a)}^2 = [(1 - R^2_2)/(N-2)] = [(1 - a^2)/(N-2)]$. Because both of these models have a single predictor the tolerance is 1.

The R^2 for a multiple (K) predictor standardized regression model can be calculated as the sum of the product between the k^{th} standardized regression coefficient ($\hat{\beta}_k$) and the bivariate correlation (r_{YX_k}) between Y and the k^{th} predictor X :

$$R^2 = \sum_{k=1}^K \hat{\beta}_k r_{YX_k} \quad (4)$$

(see Pedhazur, 1997). In matrix notation, (4) can be expressed as; $R^2 = \mathbf{B}'\mathbf{R}_{YX}$, where \mathbf{B}' is a $1 \times K$ vector of standardized sample regression coefficients and \mathbf{R}_{YX} is a $K \times 1$ vector of bivariate predictor-outcome (X - Y) correlations. For the $K = 2$ predictor Model 3, the vector of standardized sample regression coefficients is $\mathbf{B}' = [b \ t']$; the sample estimates of β and τ' , respectively.

Although one would know the bivariate correlations from sample data, Models 1 – 3 are used to demonstrate how Model 2 affects collinearity in Model 3 and as a result affects the standard error of the product tests. Model 1 defines the population bivariate correlation of X with Y as τ . And although Model 3 implies the population bivariate correlation of X with Y as $\tau = (\tau'a\beta)$, there is no direct knowledge of the population bivariate correlation of Z with Y , only the partial relationship β . Knowing that the standardized regression coefficients can be solved as: $\mathbf{B} = \mathbf{R}_{XX}^{-1}\mathbf{R}_{YX}$, where \mathbf{R}_{XX}^{-1} is the inverse of the $K \times K$ matrix of correlation among the predictors (\mathbf{R}_{XX}). The $K = 1$ predictor Model 2 defines the population correlation among the two predictors in Model 3 (X and Z) as α , and thus, the 2×2 matrix of predictor sample correlations is:

$$\mathbf{R}_{XX} = \begin{bmatrix} 1 & a \\ a & 1 \end{bmatrix} \quad (5)$$

The vector $K \times 1$ vector of bivariate predictor-outcome (X - Y) correlations can be solved as:

$$\mathbf{R}_{YX} = \mathbf{R}_{XX}\mathbf{B}. \quad (6)$$

Substituting (5) into (6),

$$\mathbf{R}_{YX} = \mathbf{R}_{XX}\mathbf{B} = \begin{bmatrix} 1 & a \\ a & 1 \end{bmatrix} \begin{bmatrix} b \\ t' \end{bmatrix} = \begin{bmatrix} b + at' \\ ab + t' \end{bmatrix}. \quad (7)$$

Then R^2 can be solved as:

$$R^2 = \mathbf{B}'\mathbf{R}_{XX}\mathbf{B} = \mathbf{B}'\mathbf{R}_{YX} = [b \ t'] \begin{bmatrix} b + at' \\ ab + t' \end{bmatrix} = [b^2 + t'ab] + [t'ab + t'^2] \quad (8)$$

Thus, the Full Model $R^2_3 = t'^2 + b^2 + 2abt'$ and $MSE_3 = (1 - R^2_3)(N-1)/(N-3)$. This model has two predictors, and thus, the standard errors for each coefficient will be affected by their tolerance (i.e., the correlation between the predictors X and Z). With a two predictor model both predictors have the same

tolerance, which for Model 3 is equal to $[1 - R^2_2] = [1 - a^2]$. The squared standard errors for both the b and t' standardized coefficients are:

$$SE^2_{(t')} = SE^2_{(b)} = (1 - R^2_3) / [(N-3)(1 - R^2_2)] = (1 - R^2_3) / [(N-3)(1 - a^2)]. \tag{9}$$

Thus, for the Sobel test, $z_S = ab / (SE_{(ab)S}) = ab / (b^2 SE^2_{(a)} + a^2 SE^2_{(b)})^{1/2}$.

$$z_S = \frac{ab}{\sqrt{\frac{b^2(1-a^2)}{(N-2)} + \frac{a^2(1-[t'^2+b^2+2abt'])}{(N-3)(1-a^2)}}}. \tag{10}$$

For the Aroian and Goodman tests, the covariance of the a and b coefficients is:

$$SE^2_{(a)}SE^2_{(b)} = \frac{(1-a^2)}{(N-2)} * \frac{(1-[t'^2+b^2+2abt'])}{(N-3)(1-a^2)} = \frac{(1-[t'^2+b^2+2abt'])}{(N-2)(N-3)}. \tag{11}$$

Therefore, the Aroian test is $z_A = ab / (SE_{(ab)A}) = ab / (b^2 SE^2_{(a)} + a^2 SE^2_{(b)} + SE^2_{(a)}SE^2_{(b)})^{1/2}$, and thus,

$$z_A = \frac{ab}{\sqrt{\frac{b^2(1-a^2)}{(N-2)} + \frac{(1-[t'^2+b^2+2abt'])}{(N-3)} [\frac{a^2}{(1-a^2)} + \frac{1}{(N-2)}]}}. \tag{12}$$

The Goodman test is:

$$z_G = \frac{ab}{\sqrt{\frac{b^2(1-a^2)}{(N-2)} + \frac{(1-[t'^2+b^2+2abt'])}{(N-3)} [\frac{a^2}{(1-a^2)} - \frac{1}{(N-2)}]}}. \tag{13}$$

With a in the numerator and a^2 and $(1 - a^2)$ in the denominator, it is apparent that these product tests are a nonlinear function of the a coefficient, and therefore, the magnitude and power of these three tests will not necessarily increase with a .

Figure 2 shows the magnitude for each test as a function of the a coefficient for a fixed sample size of $N = 200$ and partial coefficients of $b = 0.1$ and $t' = 0$ (complete mediation). The three product tests reach a maximum between $a = 0.2$ and 0.3 and then decline and converge thereafter. Figure 3 shows the statistical power of these product tests at a two-tailed $\alpha = 0.05$ significance level in the same complete mediation model. As would be expected from the test statistics in Figure 2, the Goodman test has slightly more power, which reaches a maximum of approximately 27% power at $a = 0.20$. The power functions of the Sobel and Aroian test are similar to each other and to the Goodman test. For $a > 0.50$ these power functions converge and decline towards $\alpha = 0.05$. Both Figures 2 and 3 demonstrate that magnitude and power of the product tests can decline with increases with a . Figure 3 (right panel) shows statistical power at a two-tailed $\alpha = 0.05$ for a partial mediation model with partial coefficients of $b = 0.1$ for $t' = 0.4$. These results are similar to those in

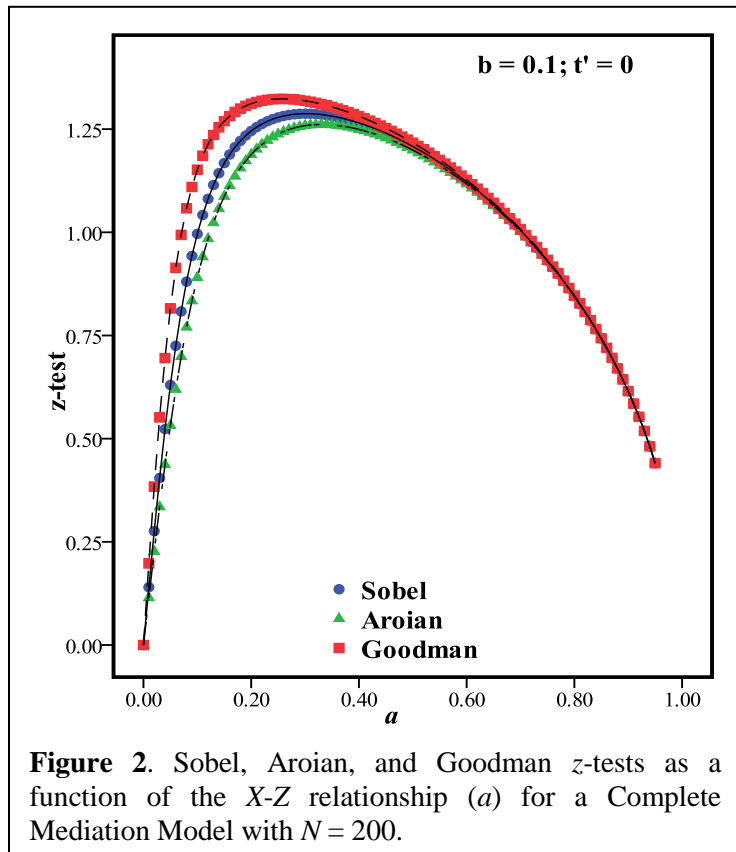


Figure 2. Sobel, Aroian, and Goodman z -tests as a function of the X - Z relationship (a) for a Complete Mediation Model with $N = 200$.

Figure 3 showing that power can decline with increases in a ; however, it also demonstrates that product tests can more powerful as tests of partial mediation. Figure 4 shows the statistical power of these product tests for a complete mediation model with a stronger Z - Y partial relationship ($b = 0.3$). As compared to Figure 3, Figure 4 shows that power the three tests are virtually identical and the point at which power reaches a maximum and then declines shifts to a larger value of $a \approx 0.45$.

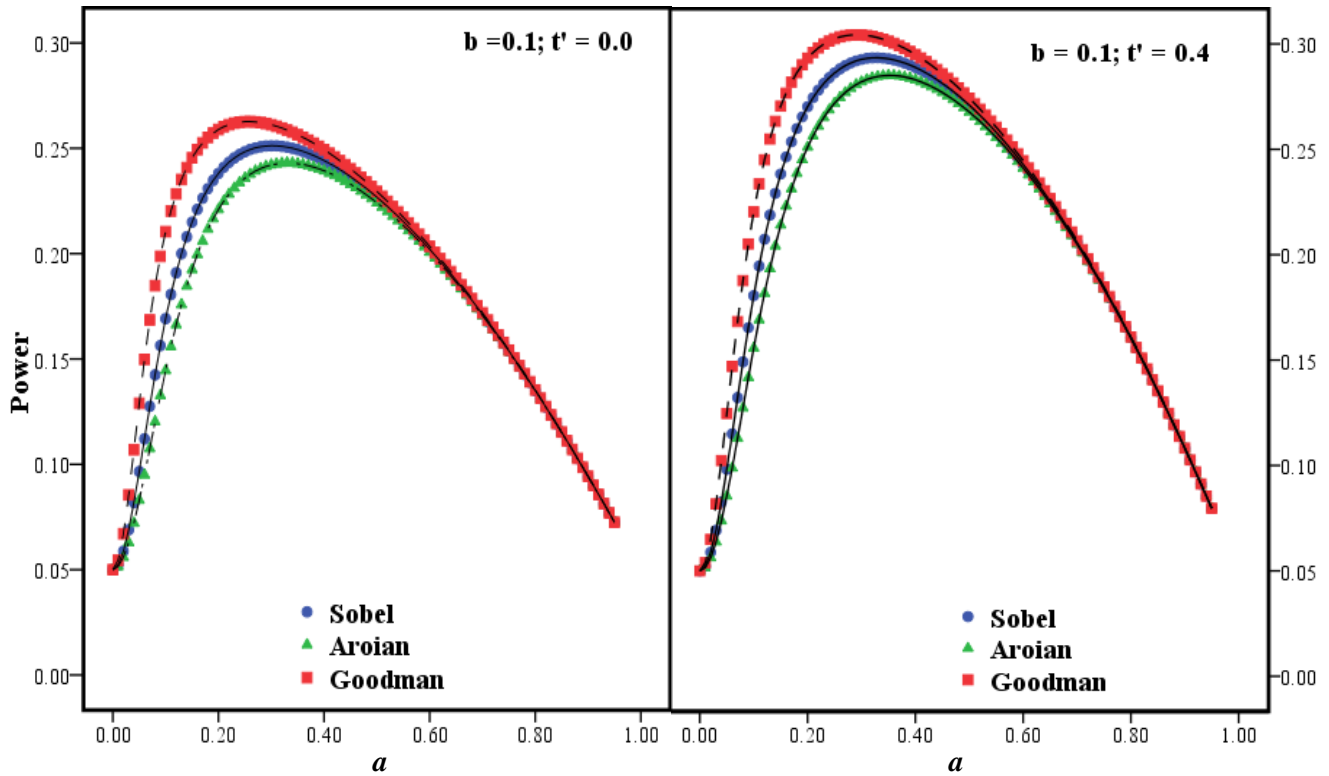


Figure 3. Power of Sobel, Aroian, and Goodman z-tests as a function of the X-Z relationship (a) for Complete (Left Panel) and Partial (Left Panel) Mediation Models at $\alpha = 0.05$ with $N = 200$.

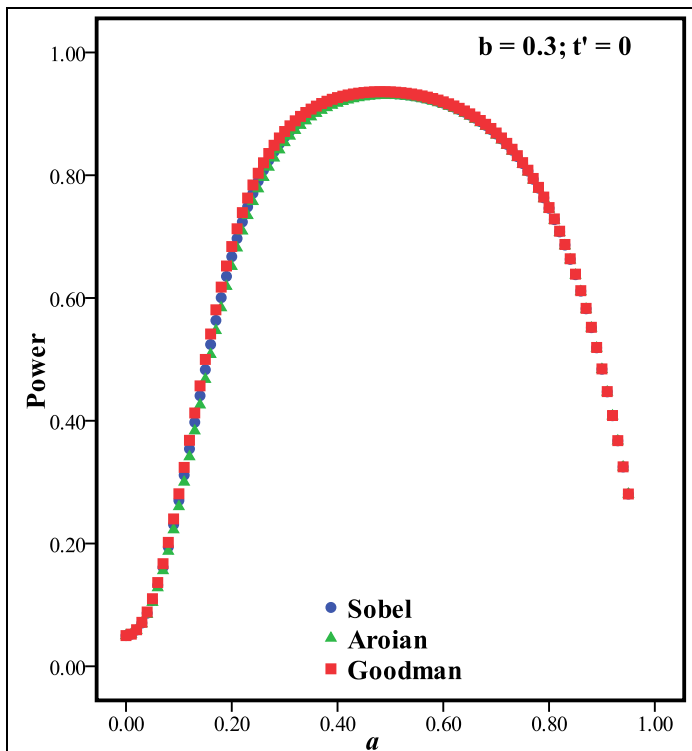


Figure 4. Power of Sobel, Aroian, and Goodman z-tests as a function a for a Complete Mediation Model at $\alpha = 0.05$ significance level with $N = 200$.

Figure 5 shows the power of the same complete mediation model as Figure 3 (Right Panel) but with a smaller sample size of $N = 25$. As compared to Figure 4, Figure 5 shows more separation between the power functions of the three product tests with the Goodman test having more power. The power of these tests reaches a maximum and then declines for a between 0.4 and 0.5, and then converge at approximately $a \approx 0.7$. Thus, it appears that the sample size (N) also affects the point at which the a coefficient begins reducing the power of the product tests. To further investigate this issue, the power of the Goodman test in a complete mediation model with a strong Z-Y partial relationship ($b = 0.85$) with sample sizes of $N = 25, 50$ and 75 was examined. Figure 6 shows that the point at which the product test reach $>99\%$ power is at $a = 0.8$ for $N = 25$, $a = 0.6$ for $N = 50$ and $a = 0.5$ for $N = 75$, which is to be expected. The point at which the power of the product test begins to decline also differs. Thus, statistical power of the product tests can decline as a function of the a coefficient; however, the sample size will have some effect on this phenomenon.

Discussion

The major conclusion is that an increase in X - Z mediational relationship (a) leads to increases in the standard error of product tests due to the necessary increase in the collinearity between the X and Z predictors in Model 3. This is because the tolerance of the predictors in Model 3 ($1 - a^2$) is a direct function of a . Although it would seem that the power of the product test would increase with a , power can actually be reduced by increasing a . In the conditions presented, but in other conditions examined but not reported, power usually reached a maximum and then began to decline with increases in a . For mediation models with stronger Z - Y partial relationship (b), the point at which power reaches a maximum tend to get shifted to a larger value of a ; however sample size will have some effect on this value.

Regardless of the τ' parameter, mediation does not occur when either the α or β (or both) parameters equal zero (i.e., mediation null hypothesis is true). For complete mediation models ($\tau' = 0$) both a and b can be large (> 0.9); however, if $a = 1$ then $b \equiv 0$ because X and Z would be completely collinear as predictors in Model 3. Yet, given the same product ab , power is a stronger function of b because increases in a also increase the standard error of the product tests. This is consistent with Hoyle and Kenny (1999) conclusion that the power of the product test of is maximal when b is somewhat larger than a .

Our findings also show that products tests are more powerful as tests of partial mediation. In partial mediation models where $\tau' \neq 0$, the sample value of t' places constraints on the values of a and b can take. That is, 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 (i.e., has all non-zero eigenvalues). Thus, holding a and b constant, for any $t' \neq 0$ that is valid (positive definite 3×3 correlation matrix), then Model 3 will explain more variation (R^2_3), which results in smaller $SE^2_{(b)}$ and $SE^2_{(ab)}$ and increased power for the product test. If t' is large then the R^2_3 will not be reduced by adding Z to Model 3, it will be at least as

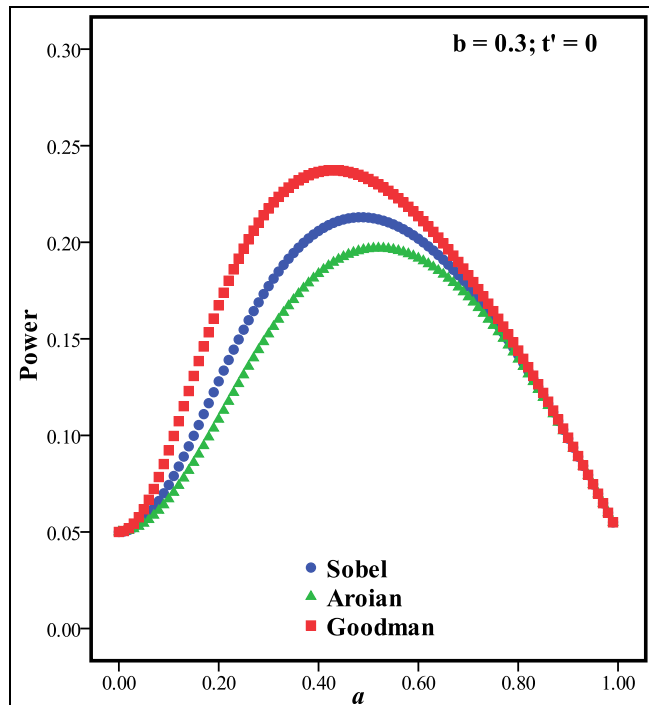


Figure 5. Power of Sobel, Aroian, and Goodman z -tests as a function of a for a Complete Mediation Model at $\alpha = 0.05$ with $N = 25$.

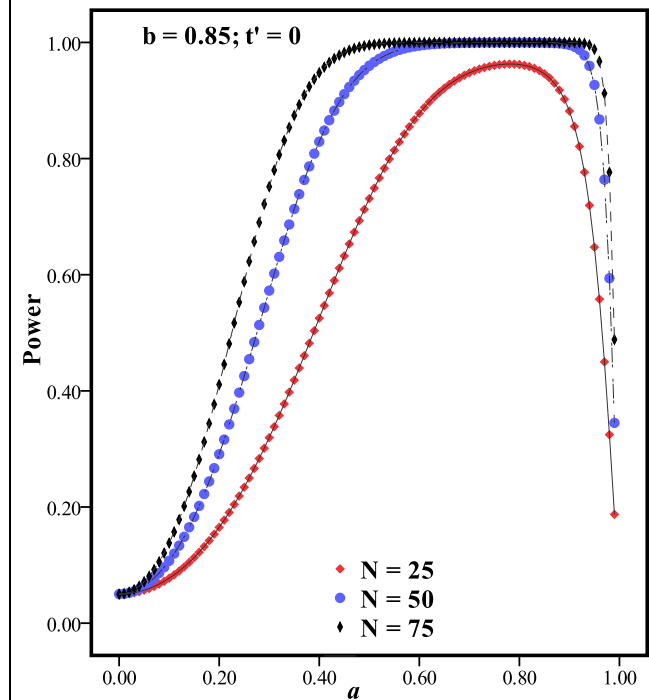


Figure 6. Power of Goodman z -test as a function of a for a Complete Mediation Model at $\alpha = 0.05$ with $N = 25, 50,$ and 75 .

large as R^2_1 ($R^2_3 \geq R^2_1$); however, a larger value of t' will place constraints on the values of a , b , and the ab product. Furthermore, although larger values of t' may still meet the criteria for partial mediation, it would seem counter to the conceptualization of mediation.

Under the conditions presented, the Goodman test demonstrated more power when the mediational relationships were smaller; however, it has been shown to have negative variance for correlation structures with a mixture of positive and negative coefficients and small samples sizes ($N < 20$), and we did not fully investigate negative values of the mediation parameters. Furthermore, the Sobel and Aroian tests have performed better than the Goodman test in more extensive simulation studies (MacKinnon et al., 1995). Also, the properties of difference tests (McGuigan & Langholtz, 1988) were not examined; however, MacKinnon et al. (1995) concluded that this test should not be used when the independent variable (X) is binary. Yet, whether one uses a product test or a difference test, the standard errors will increase as a function of the a coefficient because the product tests use $SE^2_{(b)}$ and the difference test uses $SE^2_{(t')}$ as part of the denominator. Thus, it is reasonable to expect the power of the difference test would also decline with a in many circumstances.

References

- Albert, J. M., Mediation analysis via potential outcomes models. *Statistics in Medicine*, 2008, 27, 1282 – 1304.
- Aroian, L. A. (1944/1947). The probability function of the product of two normally distributed variables. *Annals of Mathematical Statistics*, 18, 265-271.
- 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.
- Goodman, L. A. (1960). On the exact variance of products. *Journal of the American Statistical Association*, 55, 708-713.
- Hoyle, R. H., & Kenny, D. A. (1999). Statistical power and tests of mediation. In R. H. Hoyle (Ed.), *Statistical strategies for small sample research*. (pp. 195-222). Newbury Park: Sage.
- MacKinnon, DP (2008). *Introduction to Statistical Mediation Analysis*. Mahwah NJ: Erlbaum.
- MacKinnon, D. P., Lockwood, C. M., Hoffman, J. M., West, S. G., & Sheets, V. (2002). A comparison of methods to test the significance of the mediated effect. *Psychological Methods*, 7(1), 83-104.
- MacKinnon, D. P., Warsi, G., & Dwyer, J. H. (1995). A simulation study of mediated effect measures. *Multivariate Behavioral Research*, 30, 41-62.
- McGuigan K, Langholtz B. 1988. A note on testing mediation paths using ordinary least-squares regression. Unpublished note.
- Pedhazur, E. J. (1997). *Multiple regression in behavioral research*. Harcourt Brace: Orlando, FL.
- Preacher, K. J., & Hayes, A. F. (2008). Asymptotic and resampling methods for estimating and comparing indirect effects. *Behavior Research Methods*, 40, 879-891.
- Schumacker, R. E. & Lomax, R. G. (1996). *A Beginner's Guide to Structural Equation Modeling*. Mahwah NJ: Erlbaum.
- Shrout, P. E., & Bolger, N. (2002). Mediation in experimental and nonexperimental studies: New procedures and recommendations. *Psychological Methods*, 7, 422-445.
- Sobel, M. E. (1982). Asymptotic confidence intervals for indirect effects in structural equation models. In S. Leinhardt (Ed.), *Sociological Methodology 1982* (pp. 290-312). Washington DC: American Sociological Association.

Send correspondence to:

T. Mark Beasley
 University of Alabama at Birmingham
 Email: mbeasley@uab.edu
