

BRIEF REPORTS

Advances in Testing the Statistical Significance of Mediation Effects

Brent Mallinckrodt
University of Missouri—Columbia

W. Todd Abraham, Meifen Wei,
and Daniel W. Russell
Iowa State University

P. A. Frazier, A. P. Tix, and K. E. Barron (2004) highlighted a normal theory method popularized by R. M. Baron and D. A. Kenny (1986) for testing the statistical significance of indirect effects (i.e., mediator variables) in multiple regression contexts. However, simulation studies suggest that this method lacks statistical power relative to some other approaches. The authors describe an alternative developed by P. E. Shrout and N. Bolger (2002) based on bootstrap resampling methods. An example and step-by-step guide for performing bootstrap mediation analyses are provided. The test of joint significance is also briefly described as an alternative to both the normal theory and bootstrap methods. The relative advantages and disadvantages of each approach in terms of precision in estimating confidence intervals of indirect effects, Type I error, and Type II error are discussed.

Keywords: statistical mediation, bootstrap methods, test of joint significance, statistical power

Frazier, Tix, and Barron (2004) provided a valuable contribution to counseling psychologists by describing quantitative methods for testing moderator and mediator effects in multiple regression contexts. Their review presented a thorough discussion of critical methodological and conceptual issues, together with detailed guidance for implementing statistical procedures. Frazier et al. pointed out that much of the research published previously in the *Journal of Counseling Psychology* had not followed these best practices. The guidelines and checklists provided in Frazier et al.'s review will no doubt play an important role in increasing the quality of counseling psychology research. In their review, Frazier et al. devoted considerable attention to the method for testing mediating variables popularized by Kenny and associates (Baron & Kenny, 1986; Kenny, Kashy, & Bolger, 1998). Because this method relies on the assumption that indirect (i.e., mediation) effects are distributed normally, we refer to it as the *normal theory* approach. This

approach is frequently used in studies of mediation effects published in the *Journal of Counseling Psychology* and in psychological research more generally. However, Frazier et al. cited a comparison of methods reporting that the normal theory approach lacks statistical power relative to other methods (MacKinnon, Lockwood, Hoffman, West, & Sheets, 2002).

One alternative to the normal theory approach is the method described by Shrout and Bolger (2002) based on bootstrap data-resampling procedures to establish confidence intervals for testing the statistical significance of an indirect effect. The purpose of this article is to supplement Frazier et al.'s (2004) review by providing an example and a step-by-step guide for performing these bootstrap procedures. MacKinnon, Lockwood, and Williams (2004) reported that the bootstrap approach affords greater statistical power than the normal theory approach and that the advantages are relatively greater as either the sample size or the effect size decreases. Recent advances in software implementation have made the bootstrap method much more user-friendly. In an online supplement to this report (<http://dx.doi.org/10.1037/0022-0167.53.3.372.supp>), we present a step-by-step guide for conducting Shrout and Bolger's (2002) bootstrap tests of indirect effects using commonly available statistical software packages. Finally, because data simulation research suggests that the bootstrap method is more susceptible to Type I error than some alternatives (MacKinnon et al., 2004), we describe an alternative to both the normal theory and bootstrap approaches called the *test of joint significance* (Kenny et al., 1998). We conclude by discussing the relative advantages of both methods as alternatives to normal theory procedures.

Editor's Note. Joel R. Levin served as guest editor for this manuscript.—BSM

Brent Mallinckrodt, Department of Educational, School, and Counseling Psychology, University of Missouri—Columbia; W. Todd Abraham and Meifen Wei, Department of Psychology, Iowa State University; Daniel W. Russell, Department of Human Development and Family Studies and Institute for Social and Behavioral Research, Iowa State University.

Additional materials are on the Web at <http://dx.doi.org/10.1037/0022-0167.53.3.372.supp>.

We gratefully acknowledge Patricia A. Frazier, William Hoyt, Kristopher Preacher, and Katherine Daly for their comments on earlier versions of this article.

Correspondence concerning this article should be addressed to Brent Mallinckrodt, Department of Educational, School, and Counseling Psychology, 16 Hill Hall, University of Missouri, Columbia, MO 65211. E-mail: mallinckrodtb@missouri.edu

Critique of the Normal Theory Method

Frazier et al. (2004) described the basic normal theory approach as involving four component steps that we have symbolized in Figure 1. First, there must be a significant correlation between the

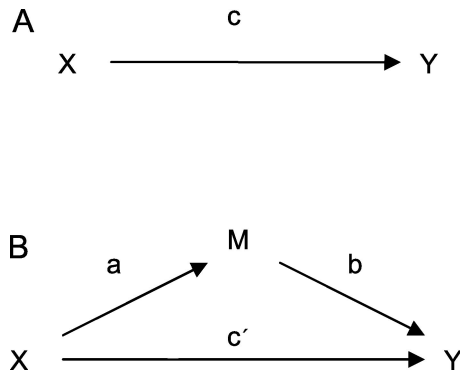


Figure 1. The mediation model requirements based on Baron and Kenny's (1986) method. A: The direct effect. B: The mediation model.

predictor variable X and the dependent variable Y , shown as Path c in Figure 1A. Second, in Figure 1B, the predictor variable X must account for a significant proportion of the variance in the mediating variable M (Path a). Third, the mediator variable M must account for a significant proportion of variance in the dependent variable Y (Path b). Fourth, the association between the predictor variable X and the dependent variable Y must be significantly less after controlling for the variance shared between the mediator and dependent variable. That is, Path c' must be significantly less than Path c . However, Kenny et al. (1998) argued that the first of these conditions is not necessary. Frazier et al. (2004) described several situations in which mediation may occur in the absence of a significant relationship between the predictor and dependent variable. Examples include suppressor variable situations in which multiple mediating variables cancel the influence of the predictor on the dependent variable (Kenny et al., 1998; MacKinnon et al., 2002) or research designs in which the predictor variable occurs much earlier in time than the dependent variable (Shrout & Bolger, 2002).

Perhaps because Baron and Kenny (1986) specifically enumerated only the first three steps described above, some researchers testing mediation have relied on satisfying these three conditions alone. MacKinnon et al. (2002) referred to the general three-step approach as the *causal steps* method. Of the three general types of methods MacKinnon and colleagues examined, variations of the causal steps approach tended to have the lowest statistical power. However, Baron and Kenny (1986) clearly indicated that additional procedures are required to examine the magnitude of the mediating effect and its statistical significance. Frazier et al. (2004) recommended the method described by Kenny et al. (1998), which requires calculating the product of Paths a and b and dividing by the standard error of this cross-product to yield a Z statistic that can be evaluated for statistical significance using probabilities corresponding to the standard normal distribution. There are two variations for the standard error used to compute the Z statistic testing an indirect effect. Baron and Kenny (1986) recommended using $\sqrt{a^2sb^2 + b^2sa^2 + sa^2sb^2}$, where a and b are the unstandardized (raw) regression coefficients derived from multiple regression analysis and sa and sb are the corresponding standard errors (see Kenny et al., 1998, for an updated description of this procedure). The formula developed earlier by Sobel (1982) differs in that it omits the final sa^2sb^2 term. Because this term

tends to be very small relative to the other two terms in the denominator of the Z statistic, the two procedures typically yield equivalent results (MacKinnon et al., 2002). For convenience, in the remainder of this article we refer to Frazier et al.'s recommended procedure as the *normal theory* (NT) approach.

Our concerns about using the NT method focus on the final step used to test statistical significance. MacKinnon, Warsi, and Dwyer (1995) observed that the product of two normally distributed variables is not itself normally distributed. MacKinnon et al. (2002) reported that the sampling distribution of the product of Paths a and b tends to be asymmetric and highly kurtotic. Thus, the corresponding Z test lacks statistical power relative to methods that attempt to correct for this asymmetry. Although MacKinnon et al. (2002) found that the NT method was superior to the sequence of three causal steps approach, their simulation analyses suggested that the NT method lacks statistical power relative to other alternatives, especially when the sample size is small or the effect size is modest (see "Second-order test," Table 6, p. 94). For example, simulated samples of 50 participants afforded a power of only .33 ($\alpha = .05$, two-tailed test) to detect a moderate-sized effect ($pr = .36$). For a small effect size ($pr = .14$), a sample of 500 participants afforded power of only .53 ($\alpha = .05$, two-tailed test). MacKinnon et al. described alternatives to the NT method that offer greater statistical power and lower Type I error rates. Although in the past, these alternatives required cumbersome computation procedures or access to specialized resources (e.g., tables of critical values), MacKinnon and colleagues recently developed software code to perform indirect effects tests on the basis of the distribution of the cross-product, making use of this approach more accessible (MacKinnon, Fritz, Williams, & Lockwood, in press). In a later study, MacKinnon et al. (2004) conducted a simulation comparing the NT method with various resampling procedures (e.g., jackknife, bootstrap, and Monte Carlo), including Shrout and Bolger's (2002) *bias-corrected* bootstrap method (described in the next section). Confidence intervals derived from the NT approach exhibited poorer coverage of the actual parameter estimates and lower statistical power to detect true nonzero effects. Of the resampling methods compared, Shrout and Bolger's bias-corrected bootstrap method performed relatively best.

Bootstrap Estimation to Determine Significance of Mediation Effects

Bootstrap methods are particularly useful for examining sampling distributions. These approaches treat the collected research sample as a "population reservoir" from which a large number of random samples are drawn with continuous replacement such that the probability of selection for any given case remains equal over every random draw. Assuming a research sample of size N , selection with replacement of Case 007 as the first member of a bootstrap sample does not influence the probability of drawing Case 007 on any subsequent draw (i.e., the selection probability remains $1/N$). Thus, a given bootstrap sample will omit some members of the original sample and include other cases multiple times. Sampling with replacement makes it possible to draw a very large number of unique samples from the population reservoir of size N . In practice, one typically draws as many as 10,000–20,000 bootstrap samples and calculates a given parameter for each sample. Variability in the distribution of the parameter estimates across

the many bootstrap samples accurately models variability in the original research sample to the degree that the original sample accurately represents the population from which it is drawn. *Empirical distributions* resulting from bootstrap analyses are especially useful for estimating the confidence interval of a given population parameter. This principle is fundamental to Shrout and Bolger's (2002) method for testing the statistical significance of an indirect effect.

The following steps are adapted from Shrout and Bolger's (2002, p. 426) description of how to conduct the bootstrap procedure:

1. Using the original data set of N cases as a population reservoir, create a bootstrap sample of N cases by random sampling with replacement.
2. Calculate a , b , and $a \times b$ based on this bootstrap sample, and save the results to a file.
3. Repeat Steps 1 and 2 a total of J times.
4. Examine the distribution of J estimates, and if $\alpha = .05$, determine the 2.5 and 97.5 percentile values of $a \times b$.

Note that because the $a \times b$ distribution is typically skewed, researchers should expect that the upper and lower bounds of the 95% confidence interval resulting from this procedure will not be equally distant from the mean value of $a \times b$. Shrout and Bolger (2002) reported analyses of simulated data suggesting that results from $J = 500$ and $J = 1,000$ bootstrap samples were quite similar. However, thanks to modern desktop computer capabilities, the effort required to run 1,000, or even 10,000, bootstrap samples is only marginally greater than the effort required to run 500 samples, so we recommend analyses based on at least 10,000 bootstrap iterations. Readers are urged to consult Shrout and Bolger (2002) for detailed instructions and a very helpful decision tree to rule out the possibility (not sufficiently appreciated by many researchers) of suppression effects that confound straightforward interpretation of mediation. In small samples (e.g., $N < 80$), a further bias correction to the confidence interval boundaries is necessary to compensate for asymmetry in the distribution of bootstrap estimates (Efron & Tibshirani, 1993).

Shrout and Bolger's (2002) method specifically addresses the asymmetric nature of the sampling distribution of the $a \times b$ product term used to conduct NT tests. MacKinnon et al.'s (2004) simulation suggests this method is a statistically powerful alternative to the NT approach. Shrout and Bolger (2002) observed that the product of two normally distributed variables with positive means forms a sampling distribution that tends to be positively skewed, whereas the product of two normally distributed variables with means that are opposite in sign is typically negatively skewed. Consequently, confidence intervals surrounding point estimates of an indirect effect (i.e., the product term) tend to be asymmetric. Shrout and Bolger reviewed results of simulation studies (MacKinnon & Dwyer, 1993; Stone & Sobel, 1990), demonstrating that the confidence interval derived from a skewed sampling distribution tends to be too wide in the direction of the null hypothesis (i.e., $a \times b = 0$) and too narrow in the direction of the alternative hypothesis (i.e., $a \times b \neq 0$). In other words, the NT method, which

assumes a symmetric distribution of the standard error around the estimated mediation effect, lacks power to detect true mediation effects. Building on Bollen and Stine's (1990) use of bootstrap methods to construct asymmetric confidence intervals around the product of two regression coefficients, Shrout and Bolger described the application of bootstrap resampling methods to construct empirical distributions for $a \times b$ cross-product values. Cross-product values corresponding to the $\alpha/2 \times 100$ and $1 - \alpha/2 \times 100$ percentile values of the empirically derived $a \times b$ distribution provide lower and upper limits, respectively, for confidence intervals of the indirect effect. Shrout and Bolger cited the work of Efron and Tibshirani (1993) in cautioning that bootstrapped sampling distributions for parameter estimates are not normally distributed. Therefore, percentile confidence intervals obtained from bootstrapped distributions will yield intervals that retain bias and need further correction.

Results from recent simulation work comparing these bootstrap methods to traditional approaches indicate that resampling methods outperform both NT methods and methods based on the distribution of cross-products. Further, MacKinnon et al. (2004) reported that the bootstrap confidence intervals adjusted for bias exhibited the highest levels of statistical power. Bias-corrected bootstrap confidence intervals for the indirect effect performed the best of the methods examined when the population indirect effect was equal to zero.

An Example Using Bootstrap Estimation

To provide a detailed illustration of how to conduct a bootstrap estimation analysis with a small sample of the type especially appropriate for this procedure, we drew a sample of 60 cases from a larger survey study of 430 students conducted by Mallinckrodt and Wei (2005). Demographic information and research procedures are available in the original report of this study. Among other instruments, survey packets contained the following three measures: (a) a measure of adult attachment, the Experiences in Close Relationship Scale (Brennan, Clark, & Shaver, 1998), which yields a subscale score of *attachment avoidance*; (b) the Toronto Alexithymia Scale (Bagby, Parker, & Taylor, 1994), which yields two subscale scores regarding problems identifying feelings and problems expressing feelings that were reverse scored and summed to form a measure of *emotional awareness*; and (c) a measure of global *psychological distress*, the Outcome Questionnaire (Lambert et al., 1996).

Attachment theory and previous research (e.g., Fraley, Davis, & Shaver, 1998) suggest that persons with high attachment avoidance tend to repress their feelings. Persons with high alexithymia experience strong negative emotional arousal, but they tend to lack the ability to label these feelings and communicate about them with others. Thus, Mallinckrodt and Wei (2005) hypothesized that attachment avoidance would be negatively associated with emotional awareness (i.e., positively associated with the two Toronto Alexithymia Scale subscales) and that it is this lack of emotional awareness that contributes to symptoms measured by the Outcome Questionnaire, such as problems in interpersonal relationships and generalized psychological distress. Thus, for purposes of illustration, we examined the model shown in Figure 2 involving (a) attachment avoidance as the independent variable, (b) emotional awareness as the mediator variable, and (c) psychological distress

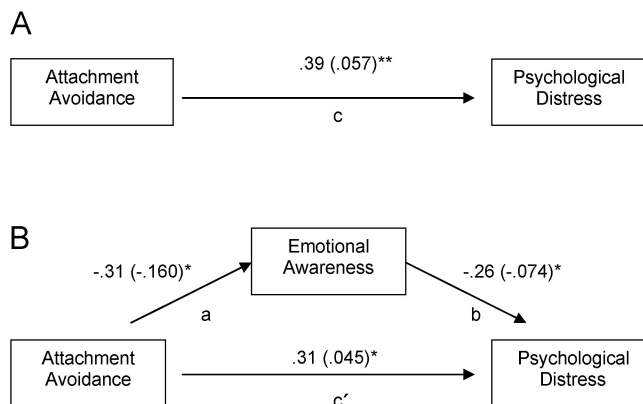


Figure 2. A three-variable mediation model. A: The direct effect model for attachment avoidance and psychological distress. B: The mediation model with emotional awareness as a mediator between attachment avoidance and psychological distress. Standardized path coefficients are shown, with corresponding unstandardized coefficients in parentheses. $N = 60$. * $p < .05$; ** $p < .01$.

as the dependent variable. The hypothesis of interest is that emotional awareness will be a significant mediator of the relationship between attachment avoidance and psychological distress.

Following the steps described by Frazier et al. (2004) for implementing the NT method, we conducted three regression analyses. In the first analysis, the dependent variable (psychological distress) was regressed on the independent variable (attachment avoidance) yielding the coefficient corresponding to Path c in Figure 2A. Standardized path coefficients appear in Figure 2, with corresponding unstandardized coefficients shown in parentheses. The unstandardized regression coefficient ($B = 0.057$) is also shown in the first line of Table 1. In the second analysis, the mediator variable (emotional awareness) was regressed on the independent variable (attachment avoidance) to obtain the regression coefficient for Path a in Figure 2B ($B = -0.160$). In the third analysis, the dependent variable (psychological distress) was regressed simultaneously on both the mediator (emotional awareness) and the independent variable (attachment avoidance). This analysis provided the unstandardized regression coefficients for

Paths b ($B = -0.074$) and c' ($B = 0.045$), respectively, in Figure 2.

Results presented in the first column of the first three rows of Table 1 indicate that the three critical conditions of the NT method were satisfied. The raw (unstandardized) regression coefficients and standard errors for Paths a and b from Table 1 provide the terms needed for the Z statistic used in the final step of the NT method. The numerator is the product of the raw regression coefficients for Paths a and b ($-0.160 \times -0.074 = 0.0118$), and the denominator is $\sqrt{a^2sb^2 + b^2sa^2 + sa^2sb^2}$, as described previously, where a and b are the unstandardized (raw) regression coefficients and sa and sb are the corresponding standard errors. Using this formula, we obtained the following standard error ($SE_{ab} = 0.0077$), Z statistic ($Z = 1.53$), and corresponding probability level ($p = .123$). We obtained similar results using Sobel's formula that omits the sa^2sb^2 term ($SE = 0.0074$, $Z = 1.60$, $p = .109$). On the basis of either method, one would conclude that the indirect effect of attachment avoidance on psychological distress, mediated by emotional awareness, is not statistically significant.

To perform the bootstrap analysis, we used the graphical interface of Amos 5.0 (Arbuckle, 2003), which is the one currently available software package we examined that directly produces bootstrapped percentile and bias-corrected confidence intervals for indirect effects. We began by drawing a three-variable path diagram similar to Figure 2B, including error terms for the endogenous mediator and dependent variables. Next, the bootstrap option was selected from the View/Set menu under the Analysis Properties submenu. We requested 1,000 bootstrap samples, drawn by default with replacement from the full data set of 60 cases. Note that a request for 95% confidence intervals must be used to override the Amos default that provides 90% confidence intervals. Although one should select only bias-corrected intervals, we selected both the bias-corrected and percentile (i.e., uncorrected) options for illustration purposes. We requested bootstrap estimates of indirect, direct, and total effects through the Output submenu. Bootstrapped estimates of the a , b , and c' path coefficients are shown in Table 1. Note that these values are essentially "means of means" (\bar{B}) and mean standard errors based on the 1,000 empirical samples. As would be expected, they differ only slightly from the results of the regression analyses.

Table 1
Illustration of Standard Normal (NT) and Bootstrap Methods to Test Significance of Mediation Effects

Path/effect	Regression result		Bootstrap estimate		95% Confidence interval		
	B	SE	\bar{B}	\overline{SE}	Standard normal NT	Bootstrap percentile	Bootstrap with bias correction
c (Avoid \rightarrow OQ)	-0.057**	0.018	—	—			
a (Avoid \rightarrow EA)	-0.160*	0.065	-0.161	0.075	-0.287, -0.033	-0.316, -0.021	-0.318, -0.024
b (EA \rightarrow OQ)	-0.074*	0.035	-0.077	0.037	-0.143, -0.005	-0.152, -0.007	-0.145, 0.000
c'	0.045*	0.018	0.044	0.020	0.010, 0.080	0.003, 0.083	0.005, 0.084
$a \times b$	0.012 ^a	0.007	0.012	0.011	-0.002, 0.026	0.0002, 0.0397 ^b	0.0004, 0.0413 ^c

Note. $N = 60$. NT = Baron-Kenny-Sobel method. Estimates are unstandardized. Avoid = attachment avoidance; EA = emotional awareness; OQ = psychological distress (as measured by the Outcome Questionnaire).

^a $p = .123$. ^b $p = .036$. ^c $p = .032$.

* $p < .05$. ** $p < .01$.

The last three columns of Table 1 show the upper and lower limits for the 95% confidence intervals calculated with the three methods. For purposes of comparison, we first present the standard normal symmetric confidence intervals derived by adding and subtracting the NT standard error estimates from the mean indirect effect (i.e., $1.96 \times SE$). The second set of confidence intervals shows results of the bootstrapped percentile method reported by Amos. These values correspond to the 2.5th and 97.5th percentiles from lowest to highest rank-ordered estimates of the indirect effect derived from the 1,000 samples. Note that although the standard normal (i.e., NT method) confidence intervals center on the mean, as expected the percentile confidence intervals do not. For example, the midpoint of the 95% confidence interval for percentile estimates of the $a \times b$ indirect effect is $(0.0002 + 0.0397)/2 = 0.02$, whereas the corresponding bootstrap mean is 0.012. The percentile confidence interval for the indirect effect is wider and extends farther from zero than the corresponding NT confidence interval. Note that on the basis of results of data simulation studies (MacKinnon & Dwyer, 1993; MacKinnon et al., 2002; Stone & Sobel, 1990, Shrout and Bolger (2002) observed that the 95% confidence interval obtained from the assumed normal sampling distribution of indirect effects tends to be too wide in the direction of the null hypothesis ($a \times b = 0$) and too narrow in the direction of the alternative hypothesis. This asymmetry is also apparent in our example findings. Note that because the percentile confidence interval does not include zero, we can conclude that the indirect effect is statistically significant at the .05 level (Amos reported the corresponding $p = .036$ for the bootstrap method without bias correction). The bias-corrected 95% confidence interval is shown in the last column of Table 1. Note that this assumed more accurate confidence interval (0.0004, 0.0413) also excludes zero and therefore also supports the conclusion that the indirect effect of attachment avoidance on psychological distress through the mediator of emotional awareness is statistically significant at the .05 level (Amos reports the corresponding $p = .032$ for the bias-corrected bootstrap method).

This example illustrates how the NT method, when used to test the significance of a mediator in a small sample ($N = 60$), can sometimes lead to a conclusion that the mediation effect is not statistically significant ($p = .12$), whereas the bias-corrected bootstrap method described by Shrout and Bolger (2002) leads to an inference that the same mediation effect is statistically significant ($p = .032$). Of course, the mixed conclusions obtained in this illustration will not always be the case, and thus one should not overgeneralize from this single example.

A Second Alternative: The Test of Joint Significance (TJS)

Although MacKinnon et al.'s (2002) simulation study did not include the bootstrap method, the authors concluded that the TJS exhibited the best balance of Type I error and statistical power of the 14 methods that were tested. The TJS is a variant of the causal steps approach discussed earlier but requires only that the path from predictor to mediator and the path from mediator to outcome must both be statistically significant (Cohen & Cohen, 1983, p. 366; Kenny et al., 1998). Performing the test is exceedingly straightforward and involves inspection of the regression results estimating the coefficients of Paths a and b in Figure 1B. If both

coefficients are statistically significant, the researcher concludes that $\alpha \neq 0$ and $\beta \neq 0$ and that there is a significant indirect effect.

Unfortunately, we could not locate any study that directly compared the TJS with the bootstrap approach. However, MacKinnon et al.'s (2004) study, which did test the bootstrap method, simulated conditions in which one of the population parameters (α or β) was equal to zero, but the other was not. Under these conditions, the bias-corrected bootstrap method exhibited elevated levels of Type I error. Of course, this vulnerability to Type I error is not a liability for the TJS method, because the null hypothesis is not rejected if either Path a or Path b is not statistically significant, no matter how large the companion path coefficient may be. Thus, although there have been no direct comparisons, there is reason to believe that the TJS may perform better than the bootstrap method with respect to Type I error, especially in circumstances in which one of the component paths of the indirect effect is not equal to zero in the population. Another point in favor of the TJS is its ease of use and straightforward interpretation. Standard multiple regression output testing the path model is all that a researcher requires. This output can also provide estimates of the effect size and corresponding confidence intervals for each component path of the mediation path diagram. However, as previously described, the product of these components tends to be a biased estimate of the magnitude of the indirect effect and corresponding confidence interval. Estimates of effect size and associated confidence interval are considered essential for a full report of research findings (Wilkinson & American Psychological Association Task Force on Statistical Inference, 1999; Vacha-Haase & Thompson, 2004).

Summary and Conclusions

We intend this article to supplement the important work of Frazier et al. (2004), whose review highlighted the NT approach to testing statistical significance of mediator variables popularized by Kenny and associates (Baron & Kenny, 1986; Kenny et al., 1998). However, recent research with simulated data suggests that the NT method lacks statistical power relative to other available approaches (MacKinnon et al., 2002). Unfortunately, in some areas of counseling psychology research (e.g., studies of diverse populations, counseling process or outcome studies), it is often quite difficult to collect samples large enough or find multiple indicators for the latent variable analyses that are clearly superior to multiple regression approaches for examining mediating variables. It is incumbent on researchers to do all they can to maximize statistical power, especially in research involving small samples. Of course, choice of statistical technique is only one of several methods researchers can use to decrease the likelihood of making a Type II error. One of the advantages of latent variable analyses is increased reliability of measurement. Hoyle and Robinson (2003) warned about the bias introduced into estimates of mediation effects by measurement error. They recommended that the mediating variable should be measured with a reliability of at least .90. Clearly this level of precision will be very difficult to achieve in many counseling psychology studies that rely on manifest rather than latent variables.

On the basis of the results of previous simulated data comparison studies, we offer the following recommendations to build on those provided by Frazier et al. (2004). The comparison of 14 methods (that did not include the bootstrap approach) conducted

by MacKinnon et al. (2002) concluded that the TJS provides the best balance of control for both Type I and Type II errors of the methods examined and in this regard is superior to the NT approach. Other simulation research (MacKinnon et al., 2004) suggests that the bootstrap method is vulnerable to Type I error when one of the two component paths of the mediation effect is zero in the population but deviates from zero in sample data due to sampling error and the other path is substantially greater than zero. Because the TJS does not share this vulnerability, it may be the method of choice for researchers who are not interested in obtaining an estimate of the magnitude of the indirect effect and are relatively more concerned about Type I error than Type II error, especially when they have reason to believe that either $\alpha \neq 0$ or $\beta \neq 0$. Note that we were unable to locate any study that has directly compared the TJS with the bootstrap method. Hence, the relative advantages of these two approaches remain to be determined.

Of the three approaches described here, only the NT and bootstrap methods provide an estimate of the magnitude of the indirect effect, test its statistical significance, and determine confidence intervals for the point estimate. Because most counseling psychology researchers are interested in these aspects of a mediation analysis, their choice will be between the bootstrap approach, the NT approach, or perhaps one of the other alternatives described by MacKinnon et al. (2002, 2004, in press). (See the PRODCLIN program available at <http://www.public.asu.edu/~davidpm/ripl/Prodclin/>.) Recent advances in computer software make Shrout and Bolger's (2002) bootstrap procedure relatively easy to perform, although the method is still somewhat more complex and cumbersome than the NT approach—especially when considering the ease of use and accessibility of online calculators for the NT method (e.g., Preacher & Leonardelli, 2001). Given the greater ease of use and familiarity of the NT approach, many counseling psychology researchers may question whether the bootstrap method offers sufficient advantages to offset these practical disadvantages. Considering for the moment only the question of statistical power and Type II error, the simulation conducted by MacKinnon et al. (2004), which modeled sample sizes of $N = 25, 50, 100,$ and 200 , suggested that the bootstrap method performed better than the NT approach at every level of sample size and effect size examined. In addition, the relative advantages in statistical power become greater as either effect size or sample size decreases. Because the bias-corrected bootstrap estimation procedure described in this article offers greater statistical power and confidence interval precision at every combination of sample size and effect size examined by MacKinnon et al. (2004), there are no circumstances in which we can recommend the NT method in preference to the bootstrap approach from the standpoint of statistical power and Type II error.

Despite the compelling advantages in terms of ease of use, we strongly caution researchers against analyzing their data first with NT methods and, only if the result is unsatisfactory, then using a bootstrap approach. Doing so inflates the experiment-wise Type I error rate in the same ways as do other forms of "data fishing." The NT method also carries a cost in terms of lost precision in estimating the confidence interval for the indirect effect. Quantitative experts have urged researchers to emphasize estimating confidence limits as a way to address the limitations of null hypothesis testing

(Kline, 2004; Krantz, 1999). Use of bootstrap methods provides greater precision when calculating confidence intervals regardless of the sample size, effect size, or level of statistical significance suggested by the NT findings.

We hope this article has provided a persuasive rationale for, and a sufficiently clear description of how to implement, Shrout and Bolger's (2002) bootstrap procedure. Our goal is to encourage counseling psychology researchers to use this approach. To assist researchers in implementing these methods, we have prepared a step-by-step description of how to obtain bootstrapped bias-corrected confidence intervals for indirect effects using commonly available statistical computing software packages (see Appendix A, available as an online supplement to this article: <http://dx.doi.org/10.1037/0022-0167.53.3.372.supp>). To facilitate comparisons across software packages, we use the data discussed here to present upper and lower limits of bootstrapped percentile and bias-corrected confidence intervals for the indirect effect of attachment avoidance on psychological distress through emotional awareness (see Appendix B, available through the same online link as above).

For researchers interested in applying these methods to test indirect effects in manifest (observed) variable models, we recommend macros (Preacher & Hayes, 2005) and an accompanying manuscript (Preacher & Hayes, 2006) that are available for download and use in either SPSS or SAS. These macros offer great flexibility in testing indirect effects allowing for the examination of complex mediation pathways with or without the inclusion of observed covariates (both sets of macro code are available at <http://www.quantpsy.org>). Specialized software packages such as Mplus (Muthén & Muthén, 2004a, 2004b) or Amos (Arbuckle, 2003) also directly produce bootstrapped bias-corrected confidence intervals for indirect effects if requested. Both Amos and Mplus allow for testing indirect effects within manifest or latent variable models. However, Mplus is currently more flexible than Amos in that one can explicitly request bias-corrected intervals for any specific indirect effect, whereas Amos only produces an interval for the total indirect effect, thereby making examination of specific mediation pathways more cumbersome.

Researchers are cautioned to use versions of Mplus (Version 3.12 or later) that incorporate a correction to an unreported problem with the computation of bias-corrected bootstrap confidence intervals (L. K. Muthén, personal communication, September 26, 2005). Other specialized structural modeling software such as EQS (Bentler, 2004a, 2004b) and LISREL (Jöreskog & Sörbom, 1996, 2003) do not currently produce bias-corrected confidence intervals for indirect effects. However, these programs do provide information necessary for computing these intervals by hand; we describe how to do so in Appendix A. We hope the steps provided in Appendix A will assist researchers who choose to implement the bootstrap method. Although no single currently available software package is ideal, the steps provided here should aid researchers in selecting a software package or adapting the steps necessary to obtain bias-corrected confidence intervals to individual needs and software availability. As software is continually upgraded, it is likely that the bootstrap procedure will become increasingly easier to perform.

References

- Arbuckle, J. L. (2003). Amos 5.0 [Computer software]. Chicago: SPSS.
- Bagby, R. M., Parker, J. D. A., & Taylor, J. G. (1994). The twenty-item Toronto Alexithymia Scale—I: Item selection and cross-validation of the factor structure. *Journal of Psychosomatic Research, 38*, 23–32.
- 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.
- Bentler, P. (2004a). EQS 6.1 [Computer software]. Encino, CA: Multivariate Software.
- Bentler, P. (2004b). *EQS structural equations program manual*. Encino, CA: Multivariate Software.
- Bollen, K. A., & Stine, R. (1990). Direct and indirect effects: Classical and bootstrap estimates of variability. *Sociological Methodology, 20*, 115–140.
- Brennan, K. A., Clark, C. L., & Shaver, P. R. (1998). Self-report measurement of adult attachment: An integrative overview. In J. A. Simpson & W. S. Rholes (Eds.), *Attachment theory and close relationships* (pp. 46–76). New York: Guilford.
- Cohen, J., & Cohen, P. (1983). *Applied multiple regression/correlation analysis for the behavioral sciences*. Hillsdale, NJ: Erlbaum.
- Efron, B., & Tibshirani, R. (1993). *An introduction to the bootstrap*. New York: Chapman & Hall/CRC.
- Fraley, R. C., Davis, K. E., & Shaver, P. R. (1998). Dismissing-avoidance and the defensive organization of emotion, cognition, and behavior. In J. A. Simpson & W. S. Rholes (Eds.), *Attachment theory and close relationships* (pp. 249–279). New York: Guilford.
- Frazier, P. A., Tix, A. P., & Barron, K. E. (2004). Testing moderator and mediator effects in counseling psychology research. *Journal of Counseling Psychology, 51*, 115–134.
- Hoyle, R. H., & Robinson, J. I. (2003). Mediated and moderated effects in social psychological research: Measurement, design, and analysis issues. In C. Sansone, C. Morf, & A. T. Panter (Eds.), *Handbook of methods in social psychology* (pp. 213–233). Thousand Oaks, CA: Sage.
- Jöreskog, K. G., & Sörbom, D. (1996). *LISREL 8: User's reference guide*. Chicago: Scientific Software.
- Jöreskog, K. G., & Sörbom, D. (2003). LISREL 8.54 [Computer software]. Chicago: Scientific Software.
- Kenny, D. A., Kashy, D. A., & Bolger, N. (1998). Data analysis in social psychology. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *The handbook of social psychology* (4th ed., pp. 233–265). New York: Oxford University Press.
- Kline, R. B. (2004). *Beyond significance testing: Reforming data analysis methods in behavioral research*. Washington, DC: American Psychological Association.
- Krantz, D. H. (1999). The null hypothesis testing controversy in psychology. *Journal of the American Statistical Association, 94*, 1372–1381.
- Lambert, M. J., Hansen, N. B., Umphress, V., Lunnen, K., Okiishi, J., Burlingame, G. M., & Reisinger, C. W. (1996). *Administration and scoring manual for the Dependent Questionnaire (OQ-45.2)*. Stevenson, MD: American Professional Credentialing Services.
- MacKinnon, D. P., & Dwyer, J. H. (1993). Estimating mediated effects in prevention studies. *Evaluation Review, 17*, 144–158.
- MacKinnon, D. P., Fritz, M. S., Williams, J., & Lockwood, C. M. (in press). Distribution of the product confidence limits for the indirect effect program PRODCLIN. *Behavioral Research Methods*.
- MacKinnon, D. P., Lockwood, C. M., Hoffman, J. M., West, S. G., & Sheets, V. (2002). A comparison of methods to test mediation and other intervening variable effects. *Psychological Methods, 7*, 83–104.
- MacKinnon, D. P., Lockwood, C. M., & Williams, J. (2004). Confidence limits for the indirect effect: Distribution of the product and resampling methods. *Multivariate Behavioral Research, 39*, 99–128.
- MacKinnon, D. P., Warsi, G., & Dwyer, J. H. (1995). A simulation study of mediated effect measures. *Multivariate Behavioral Research, 30*, 41–62.
- Mallinckrodt, B., & Wei, M. (2005). Attachment, social competencies, interpersonal problems, and psychological distress. *Journal of Counseling Psychology, 52*, 358–367.
- Muthén, B. O., & Muthén, L. K. (2004a). Mplus 3.13 [Computer software]. Los Angeles: Muthén & Muthén.
- Muthén, L. K., & Muthén, B. O. (2004b). *Mplus user's guide* (3rd ed.). Los Angeles: Muthén & Muthén.
- Preacher, K. J., & Hayes, A. F. (2005, December). SPSS and SAS macros for estimating and comparing indirect effects in multiple mediator models [Computer software]. Retrieved May 1, 2006, from <http://www.quantpsy.org>
- Preacher, K. J., & Hayes, A. F. (2006). Asymptotic and resampling strategies for assessing and comparing indirect effects in simple and multiple mediator models. Unpublished manuscript. Retrieved May 1, 2006, from <http://www.quantpsy.org>
- Preacher, K. J., & Leonardelli, G. J. (2001, March). Calculation for the Sobel test: An interactive calculation tool for mediation tests [Computer software]. Retrieved July 14, 2005, from <http://www.unc.edu/~preacher/sobel/sobel.htm>
- Shrout, P. E., & Bolger, N. (2002). Mediation in experimental and non-experimental 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* (pp. 290–312). Washington, DC: American Sociological Association.
- Stone, C. A., & Sobel, M. E. (1990). The robustness of estimates of total indirect effects in covariance structure models estimated by maximum likelihood. *Psychometrika, 55*, 337–352.
- Vacha-Haase, T., & Thompson, B. (2004). How to estimate and interpret various effect sizes. *Journal of Counseling Psychology, 53*, 473–481.
- Wilkinson, L., & American Psychological Association Task Force on Statistical Inference. (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist, 54*, 594–604.

Received March 1, 2005

Revision received May 9, 2006

Accepted May 9, 2006 ■