Mediation in Experimental and Nonexperimental Studies: New Procedures and Recommendations

Patrick E. Shrout and Niall Bolger
New York University

Mediation is said to occur when a causal effect of some variable *X* on an outcome *Y* is explained by some intervening variable *M*. The authors recommend that with small to moderate samples, bootstrap methods (B. Efron & R. Tibshirani, 1993) be used to assess mediation. Bootstrap tests are powerful because they detect that the sampling distribution of the mediated effect is skewed away from 0. They argue that R. M. Baron and D. A. Kenny's (1986) recommendation of first testing the $X \rightarrow Y$ association for statistical significance should not be a requirement when there is a priori belief that the effect size is small or suppression is a possibility. Empirical examples and computer setups for bootstrap analyses are provided.

Mediation models of psychological processes are popular because they allow interesting associations to be decomposed into components that reveal possible causal mechanisms. These models are useful for theory development and testing as well as for the identification of possible points of intervention in applied work.

Mediation is equally of interest to experimental psychologists as it is to those who study naturally occurring processes through nonexperimental studies. For example, social–cognitive psychologists are interested in showing that the effects of cognitive priming on attitude change are mediated by the accessibility of certain beliefs (Eagly & Chaiken, 1993). Developmental psychologists use longitudinal methods to study how parental unemployment can have adverse effects on child behavior through its intervening effect on quality of parenting (Conger et al., 1990). Mediation analysis is also used in organizational research (e.g., Marks, Zaccaro, & Mathieu, 2000; Mathieu, Heffner, Goodwin, Salas, & Cannon-Bowers, 2000), clinical research (e.g., Nolen-Hoeksema & Jackson, 2001), and prevention research (MacKinnon & Dwyer, 1993).

When researchers are able to show that the putative intervening variables account for the association of the distal variable with the outcome, they sometimes say that they have "explained" the association. Although statistical analyses of association cannot establish causal links definitively, they can provide evidence that one mediation pattern is more plausible than another, and they can provide invaluable information for the design of fully experimental studies of causal processes (see Bollen, 1989; MacKinnon, Lockwood, Hoffman, West, & Sheets, 2002).

Statistical approaches to the analysis of mediation have been discussed extensively in the psychological literature (Baron & Kenny, 1986; Collins, Graham, & Flaherty, 1998; James & Brett, 1984; Judd & Kenny, 1981; Kenny, Kashy, & Bolger, 1998; Maccorquodale & Meehl, 1948; MacKinnon & Dwyer, 1993; MacKinnon, Krull, & Lockwood, 2000; MacKinnon, Warsi, & Dwyer, 1995; Rozeboom, 1956; Sobel, 1982). In this article, we focus on the approach to mediation analysis that was articulated by Kenny and his colleagues (Baron & Kenny, 1986; Judd & Kenny, 1981; Kenny et al., 1998). As MacKinnon et al. (2002) have documented, this formulation of the mediation process has been very influential, but we acknowledge that other important frameworks exist (e.g., Collins et al., 1998).[1] We point out that new

Correspondence concerning this article should be addressed to Patrick E. Shrout, Department of Psychology, New York University, 6 Washington Place, Room 308, New York, New York 10003. E-mail: pat.shrout@nyu.edu

[1] One of the earliest discussions of mediation in psychology was initiated by Maccorquodale and Meehl (1948) and
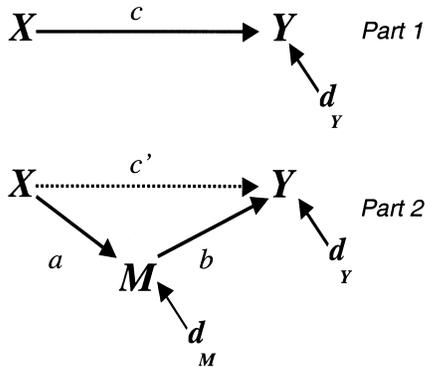
Figure 1. Path models showing total effect (Part 1) and mediated effect (Part 2) of *X* on *Y*. When mediation occurs, the *c'* path in Part 2 is smaller than the *c* path in Part 1, as indicated by dashed lines. Residual terms are displayed as *d* effects.

statistical methods that are available for analysis of mediated effects (Bollen & Stine, 1990; Lockwood & MacKinnon, 1998) have implications for Kenny's approach to statistical mediation, and we show how these new methods can be applied with relative ease to data from both experimental and nonexperimental studies.

Baron and Kenny's (1986) Approach to Statistical Mediation Analysis

An influential description of how mediation can be detected statistically was given by Baron and Kenny (1986); more recently, Kenny et al. (1998) presented an updated account. Figure 1 shows the elements of the mediation analysis. Let variable *X* be a presumed cause of variable *Y.* Variable *X* may be experimentally manipulated, or it might vary naturally in the population. When *X* is an experimental variable in a randomized study, its causal relation to *Y* can be interpreted unambiguously (Holland, 1986). When *X* is not manipulated, definitive causal inferences cannot be made about the relation of *X* to *Y.* Weaker inferences about the plausibility of causal relations can be made if certain assumptions can be shown to hold.[2] It is important to note that these assumptions are not met

refined by Rozeboom (1956), but their formulation does not lead to a strong analysis of causation such as that presented by Pearl (2000) and hence is not be considered in detail here.

simply by ordering the variables in time, as there are other possible interpretations of correlation patterns from longitudinal data, including spuriousness (e.g., Link & Shrout, 1992). When Part 1 of Figure 1 is to be used with nonexperimental data, we need a clear theoretical rationale for the possible causal relation of *X* to *Y,* and it is the development of this rationale that often brings nonexperimental researchers to mediation analysis.

For both experimental and nonexperimental data, Part 1 of Figure 1 implies that a unit change in *X* is associated with a change of *c* units in *Y* when only *X* and *Y* are considered. For example, if *Y* is a test of achievement of some technical material on a 100-point scale and *X* is a measure of the number of minutes spent studying a relevant technical manual, then *c* is the number of mastery points that one expects to improve for each minute of study.

Part 2 of Figure 1 shows a model that includes variable *M,* the proposed mediator. As noted earlier, we explicitly assume that *M* precedes *Y* in time and that it is a plausible causal agent for *Y.*[3] The mediation model assumes that *M* is affected by changes in *X;* one unit change in *X* is associated with a change of *a* units in *M.* The model also assumes that changes in *M* are associated with changes in *Y,* above and beyond the direct effect of *X* on *Y.* A unit change of *M* is associated with a change of *b* units in *Y* when *X* is held constant. As a result, *X* is said to have an indirect effect on *Y* through the mediator *M.* The size of the indirect effect is simply the product of the *X*-to-*M* and *M*-to-*Y* effects, that is, $a \times b$ (see MacKinnon et al., 1995). Continuing with the example discussed earlier, if *M* is a count of technical vocabulary and *Y* and *X* are measures of general achievement and time spent studying in minutes, then we could determine (that *a* new words are learned for each minute of study and that *b* achievement points are gained for each new word that is learned. Therefore, $a \times b$ achievement

_____

[2] Bollen (1989, pp. 40–79) summarized the assumptions as (a) association (*X* and *Y* are related), (b) direction (*X* is prior to *Y,* not vice versa), and (c) isolation (the association between *X* and *Y* is not due to another variable or process). A more detailed analysis of causation is provided by Pearl (2000).

[3] Note that even if *X* is experimentally manipulated, *M* is not. For this reason the usual caveats and concerns about causal inference apply (see Kenny et al. [1998] for a discussion of this issue).

points are gained through vocabulary improvement for each minute of study.

When the indirect effect, $a \times b$, equals the total effect, c, we say the effect of X on Y is completely mediated by M. In this case, there is no direct effect of X on Y, and the path c' in Figure 1 is equal to zero. When the indirect effect does not equal the total effect but is smaller and of the same sign, we say the effect of X on Y is partially mediated by M. In this case, the path c' is a value other than zero. In their approach to mediation analysis, Kenny and his colleagues (Baron & Kenny, 1986; Judd & Kenny, 1981; Kenny et al., 1998) did not assume that c' is zero. Their analysis proceeded with the initial assumption of some form of a partial mediation model.[4] We adopted that approach in this article.

In the restatement of Judd and Kenny's (1981) and Baron and Kenny's (1986) guidelines, for assessing mediation, Kenny et al. (1998) described four steps that should be taken. The first step is to show that X is related to Y by estimating and testing the regression coefficient c in Part 1 of Figure 1 (where Y is the outcome and X is the explanatory variable). According to Kenny et al. (1998), this first step "establishes that there is an effect that may be mediated" (p. 259). The second step is to show that X is related to M by estimating the coefficient a in Part 2 of Figure 1 (where M is the outcome and X is the explanatory variable). The third step is to show that M is related to Y while X is held constant. This step involves estimating b in Part 2 of Figure 1 in a multiple regression equation with Y as the outcome and both X and M as explanatory variables. The final step is to estimate and test the path c' in Part 2 of Figure 1 to determine if the data are consistent with complete mediation. If the data suggest that c is nonzero but its analogue c' in the multiple regression does not differ from zero, then Kenny et al. would conclude that complete mediation has occurred.

Although Baron and Kenny's (1986) recommendations in their original and restated form are influential and widely cited, they are not without their critics. Several people have questioned the necessity of testing the overall association in Step 1 (Collins et al., 1998; MacKinnon, 2000; MacKinnon et al., 2000). Still others have pointed out that the recommendations are nested in a framework that assumes that mediation processes can be analyzed by linear regression analysis (Collins et al., 1998). One could also take issue with the default use of a partial mediation model approach in cases when a complete mediation

model is hypothesized (Rozeboom, 1956). We return to possible limitations of the approach of Kenny et al. (1998) later, but we use the recommendations as a foundation for our current discussion.

When mediation has occurred, we expect that the indirect effect, $a \times b$, should be nonzero (see MacKinnon et al., 1995). However, the estimate of that effect (the product of $\hat{a}$ and $\hat{b}$, which are the sample estimates of a and b) is subject to estimation error. Baron and Kenny (1986) reported that the standard error of the indirect effect estimate can be calculated using a large-sample test provided by Sobel (1982, 1986). Variations of this formula are used to calculate the standard error of mediated effects in many popular programs for structural equation modeling (e.g., Arbuckle, 1999; Bentler, 1997; Jöreskog & Sörbom, 1993).

Until recently, researchers wishing to test the significance of indirect effects had little option but to use Sobel's (1982) large-sample test. Now, however, developments in statistical theory provide alternative methods for testing direct and indirect effects in mediation models. One particularly useful approach is the bootstrap framework, which can be applied even when sample sizes are moderate or small, that is, in the range of 20–80 cases (Efron & Tibshirani, 1993). In a groundbreaking article, Bollen and Stine (1990) showed that bootstrap methodology could be very useful in studying the sampling variability of estimates of indirect effects in mediation models. Although the bootstrap is well-known to statisticians and has been incorporated as an option in structural equation modeling programs such as EQS (Bentler, 1997) and AMOS (Arbuckle, 1999), it has only recently begun to appear in the general psychology literature (Chan, Ho, Leung, Chan, & Yung, 1999; Efron, 1988; Lee & Rodgers, 1998). We hope to encourage its widespread use in mediation analyses by describing its benefits, showing how it can make a difference in the interpretation of mediation studies, and provid-

---

[4] Persons familiar with structural equation methods will recognize that assuming a complete mediation model has important statistical and conceptual advantages. When c' can be fixed to zero, variable X is said to be an instrumental variable for the $M \rightarrow Y$ path. Instrumental variables can be used in structural equation methods to study model misspecification issues and reciprocal causal paths (Bollen, 1989). Estimates of coefficients a and b are more efficient when the complete mediation model can be assumed.

ing examples of how the new methods can be implemented using commercial structural equation software.

In the next section, we review issues in the estimation and testing of indirect and direct paths in mediation analyses. We then describe the bootstrap and illustrate the methodology with two numerical examples. In interpreting mediation in these examples, we stress the importance of confidence intervals rather than significance tests. We then explore how the bootstrap methods can be used to evaluate partial mediation effects, and we take into consideration suppression or inconsistent mediation effects (MacKinnon et al., 2000). We conclude the article with reflections on Kenny's recommendations for carrying out a regression-based mediation analysis.

Estimating and Testing Direct and Indirect Paths

We begin by reviewing some of the standard methods for estimating direct and indirect paths in Figure 1. These standard methods can be implemented in ordinary regression software programs as well as structural equation modeling software. We then review the usefulness of bootstrap methods for supplementing these methods.

In the review, we assume that X, Y, and M are variables in a representative sample of N persons from a specific population. We consider two examples of this general problem. In the first example, X, Y, and M are simply measured rather than manipulated. Instances of this type arise in cross-sectional or panel studies. We explicitly assume that a case can be made in these applications that X precedes M and M precedes Y temporally and that plausible causal relations exist between X and M and between M and Y. We also assume at this point that X and M are measured without error, but we discuss later how unreliable measures can be handled with structural equation methods. In the second example, X represents a manipulated variable in a two-group experiment. It takes the value 1 in the treatment group and 0 in the control group, and group membership is randomly assigned to participants.

For both kinds of examples, ordinary least squares regression methods[5] can be used to estimate c', a, b, and c in Figure 1. We call these estimates $\hat{c}'$, $\hat{a}$, $\hat{b}$, and $\hat{c}$. The standard errors of these estimates can be computed using standard methods for linear models

(Cohen & Cohen, 1983; Draper & Smith, 1998).[6] It is well-known from the central limit theorem that the estimates of c', a, b, and c tend to be normally distributed, and it is possible to use their estimated standard errors to construct confidence intervals and to carry out statistical tests using normal theory. Specifically, if the residuals from the regression are normally distributed and the standard error of the estimate is estimated on v degrees of freedom, then the parameter is tested with a t test on v degrees of freedom and the 95% confidence interval for a given effect (say, c') can be constructed as

$$\hat{c}' \pm s_{\hat{c}'} t_{v,.975}, \qquad (1)$$

where $S_{\hat{c}'}$ is the estimated standard error of $\hat{c}'$ and $t_{v,.975}$ is the constant from the t distribution on v degrees of freedom that marks 2.5% from each tail such that the interval contains the central symmetric 95% region of the distribution. The interval defined by Equation 1 is symmetric, and it will include zero when the two-tailed t test of $\hat{c}'$ is not significant at the .05 level. For degrees of freedom over 60, $t_{v,.975}$ is between 1.96 and 2.00 for the 95% coverage that is of usual interest.

When inferences about the indirect effects are desired, the statistical methods are not as simple. Although standard statistical theory provides expressions of the standard errors of the estimates of $\hat{a}$ and $\hat{b}$, it does not provide an estimate for the product of these two quantities. Sobel (1982, 1986) derived approximate expressions for that standard error by using a Taylor series approach, and he assumed that the product would tend to be normally distributed as the sample size gets large. Several versions of the approximate standard error have been reported (MacKinnon et al., 1995, 2002), and two of these versions are most often used in the literature:

---

[5] These are the methods implemented by default in SPSS (SPSS Inc., 1999), Excel (Microsoft Inc., 1999), and other regression software systems.

[6] The standard error formulas depend on the variability of the explanatory variables, the sample size, and the unexplained variance of the outcomes. In general, if the spread of the predictors is increased, the sample size is increased, or the residual variance of the outcome is decreased, then the standard error of the estimate of the direct effects is decreased. The stability of the estimate of the standard error is indexed by degrees of freedom associated with the residual variance.

$$s_{\hat{a}\hat{b}}(1) = \sqrt{\hat{a}^2 s_{\hat{b}}^2 + \hat{b}^2 s_{\hat{a}}^2} \qquad (2a)$$

$$s_{\hat{a}\hat{b}}(2) = \sqrt{\hat{a}^2 s_{\hat{b}}^2 + \hat{b}^2 s_{\hat{a}}^2 + s_{\hat{a}}^2 s_{\hat{b}}^2}. \qquad (2b)$$

Baron and Kenny (1986) made Equation 2b well-known, but for large samples Equation 2a is nearly the same magnitude and has been recommended (MacKinnon et al., 1995). Stone and Sobel (1990), MacKinnon and Dwyer (1993), and MacKinnon et al. (1995, 2002) have shown through simulation studies that for large samples (>200) the estimators 2a and 2b are close to the parameter they estimate. These standard error estimates are often used to test whether the indirect effect is different from zero through z statistics[7] and to construct confidence intervals. For example, the 95% confidence interval for the indirect effect would be

$$(\hat{a} \times \hat{b}) \pm s_{\hat{a}\hat{b}} z_{.975}, \qquad (3)$$

where $z_{.975}$ is equal to the constant 1.96. Even if the estimate of the standard error is accurate, this confidence bound will accurately cover 95% of the possible population values only if the product $\hat{a} \times \hat{b}$ is normally distributed. There are reasons to suspect that this assumption does not hold when the null hypothesis that $a \times b = 0$ is false, that is, when mediation is present. Although the sum of two normal random variables will have a normal distribution, their product will not be normal. Indeed, products of normal variables with positive means will tend to have a positive skew, and products of normal variables with means of opposite signs will tend to have a negative skew (Bollen & Stine, 1990; Lockwood & MacKinnon, 1998; MacKinnon et al., 2002). As we discuss below, ignoring this skew in the distribution (as Sobel's test does) can reduce the power to detect mediation when it exists in the population.

Bootstrap Distributions of Effects

Bollen and Stine (1990) explored the distribution of $\hat{a} \times \hat{b}$ in several small-sample examples using bootstrap methods (Efron & Tibshirani, 1993) and found that the positive skew was notable. Stone and Sobel (1990) examined the distribution of indirect effect estimates using simulation methods and also found evidence of skew for sample sizes less than 400. Mac-

Kinnon and Dwyer (1993) obtained similar results in their simulation studies of indirect effects. MacKinnon and Lockwood (2001) discussed the skew from the perspective of formal mathematical distribution theory. Both Stone and Sobel and MacKinnon and Dwyer observed that symmetric confidence intervals constructed using assumptions of normality tended to give asymmetric error rates. The intervals appeared to be too wide in the direction of the null hypothesis but too narrow in the direction of the alternative hypothesis. The implication of these findings is that the usual test of the indirect effect will lack statistical power to reject the null hypothesis that $a \times b = 0$. This implication has been documented in computer simulations presented by MacKinnon and his colleagues (see, e.g., MacKinnon et al., 2002).

Bollen and Stine (1990) proposed computing nonsymmetric confidence bounds for $a \times b$ using bootstrap information. The bootstrap approach allows the distribution of $\hat{a} \times \hat{b}$ to be examined empirically; the percentile confidence interval is defined simply by the cutpoints that exclude $(\alpha/2) \times 100\%$ of the values from each tail of the empirical distribution. Efron and Tibshirani (1993) provided a clear rationale for this simple percentile interval as well as some refinements that we discuss later in this article and in Appendix A.

To be concrete about how to compute the percentile confidence interval, suppose that the models in Figure 1 are studied with a sample of $N = 80$ persons. The bootstrap distribution of $\hat{a} \times \hat{b}$ is constructed by the following steps:

1. Using the original data set as a population reservoir, create a pseudo (bootstrap) sample of N persons by randomly sampling observations with replacement from the data set.

2. For each bootstrap sample, estimate $\hat{a} \times \hat{b}$ and save the result to a file.

3. Repeat Steps 1 and 2 a total of J times.

4. Examine the distribution of the estimates and determine the $(\alpha/2) \times 100\%$ and $(1 - \alpha/2) \times 100\%$ percentiles of the distribution.

The standard deviation of the $\hat{a} \times \hat{b}$ estimates also

---

[7] Note that the standard error estimates are for large samples; therefore, z rather than t tests are used, and degrees of freedom are not relevant.

Table 1
Illustration of Standard and Bootstrap Methods for Mediation Example (Simulated) With N = 80

| Effect | Estimate | SE | 95% CI | | Bootstrap | |
			Standard normal	Bootstrap percentile	M	SD
a	0.331	0.147	(0.04, 0.62)	(0.05, 0.60)	0.325	0.138
b	0.275	0.103	(0.07, 0.48)	(0.11, 0.43)	0.274	0.082
c′	−0.027	0.137	(−0.30, 0.25)	(−0.27, 0.22)	−0.029	0.122
a × b	0.091	0.053	(−0.01, 0.19)	(0.01, 0.20)	0.089	0.047
c	0.064	0.138	(−0.21, 0.34)			

Note. Data were sampled from a simulated population in which $a = 0.4$, $b = 0.3$, and $c' = 0.0$. Normal 95% CIs are computed using estimate $\pm 2 \times SE$. Percentile 95% CI for bootstrap is defined using the values that mark the upper and lower 2.5% of the bootstrap distribution. CI = confidence interval.

provides an estimate of the standard error, which could be used in the usual normal-theory confidence interval. As Lockwood and MacKinnon (1998) have pointed out, the bootstrap estimate of the standard error is similar in magnitude to estimates provided by Equations 2a and 2b.

Table 1 shows a numerical example that illustrates the usefulness of the bootstrap procedure. The example makes use of a sample of size $N = 80$ with measures of *X, M,* and *Y* from Figure 1. In this example, the population parameters in Figure 1 were set to be $a = 0.40$, $b = 0.30$, and $c' = 0.0$. Thus, in this hypothetical population the total effect, *c,* is completely mediated. Table 1 shows the results of an analysis of the $N = 80$ data set from this population.[8] Note that the standard least squares estimates of *a, b,* and *c′* for the original sample deviate somewhat from the (in this case) known population values. Table 1 also shows the means and standard deviations of the estimates from the $J = 1,000$ bootstrap samples constructed in the way described in the preceding paragraph.

The estimates (and standard errors) of the parameters from the original sample are $\hat{a} = 0.33$ (0.15), $\hat{b} = 0.28$ (0.10), and $\hat{c}' = -0.03$ (0.14). The confidence intervals in Table 1 exclude zero for *a* and *b,* which means that they are statistically significant by conventional standards. The means (and standard deviations) of the bootstrap estimates are $a = 0.3$ (0.14), $\hat{b} = 0.27$ (0.08), and $\hat{c}' = 0.03$ (0.12). The usual estimate of the indirect effect from *X* to *Y* is $\hat{a} \times \hat{b} = 0.33 \times 0.28 = 0.09$. Equation 2b from Baron and Kenny (1986) suggests that the standard error of the estimated indirect effect is approximately 0.06, and Equation 2a yields a similar value of 0.053. The usual 95% confidence interval around the indirect effect using the smaller standard error is (−0.01, 0.19). On the basis of the standard analysis, we would conclude

that the indirect effect could be zero, even though both $\hat{a}$ and $\hat{b}$ are statistically different from zero.

A different conclusion is reached when we look at the distribution of $\hat{a} \times \hat{b}$ from the bootstrap procedure.[9] The mean of the bootstrap distribution is .089, and the standard deviation is .05. Both of these values are consistent with the results of the initial analysis. However, Figure 2 shows the actual distribution of the estimated indirect effects from 1,000 bootstrap samples. From Figure 2, it is evident that the 95% confidence interval for the indirect effect should not be estimated using normal theory. The distribution has a definite positive skew, and very few of the samples led to estimates of the indirect effect that were negative or zero. The extent of the skew can be seen by comparing the actual distribution to a normal distribution with the same mean and standard deviation that is superimposed. From the bootstrap percentile confidence interval shown in Table 1, we can see that 95% of the bootstrap estimates were between the values 0.01 and 0.20. This confidence interval leads us to conclude that the indirect effect of *X* on *Y* is

---

[8] This data set was artificially constructed using simulation. Simulation was done using EQS (Bentler, 1997) software. The software essentially samples observations from a population in which the relations among the variables are characterized by parameters $a = .4$, $b = .3$, and $c' = 0$. Variables *X, M,* and *Y* all had variances of 1.0 in this hypothetical population, and so the parameters can be interpreted as standardized coefficients.

[9] The bootstrap procedure was implemented using options in EQS. The regression effects were estimated using unweighted least squares, and 100% of the bootstrap samples converged. Details of computation issues are discussed later in this article.
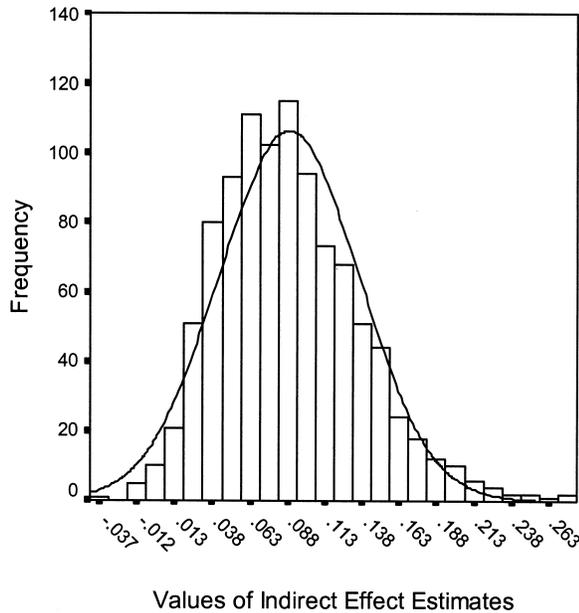
Figure 2. Bootstrap distribution of indirect effect ($\hat{a} \times \hat{b}$). $SD = .05$; $M = .089$; $N = 1,000$.

Figure 3. Bootstrap distribution of direct effect c'. $SD = .12$; $M = -.029$; $N = 1,000$.

significantly different from zero.[10] This result not only contrasts with the normal-theory confidence interval of Equation 3 but also contradicts the initial test of an association between *X* and *Y*. In this sample of 80 observations, the correlation between *X* and *Y* is only .06, and this number is not significantly different from zero when tested as a simple bivariate association.

It is not always the case that inferences based on bootstrap procedures will differ from the standard normal-theory method. Table 1 shows the 95% confidence bounds for the regression coefficients *a, b,* and *c'* as well as for the indirect effect. The substantive conclusions from these bounds are similar to those obtained using the standard method. Consider the result for the direct effect coefficient *c',* for example. This is the direct effect of *X* on *Y* when the mediator variable *M* is held constant. Recall that the original sample of 80 was simulated from a population in which $c' = 0$, and the sample result was $\hat{c}' = -.03$. The normal-theory confidence interval based on Equation 1 is $-.03 \pm (.14)(1.99) = (-.30, .25)$, where .14 is the estimated standard error of $\hat{c}'$ and 1.99 is the value of the *t* distribution on 77 degrees of freedom that yields a 95% confidence interval.

The bootstrap procedure yields a similar conclusion. Figure 3 shows the shape of the bootstrap distribution for this parameter. When the estimate was
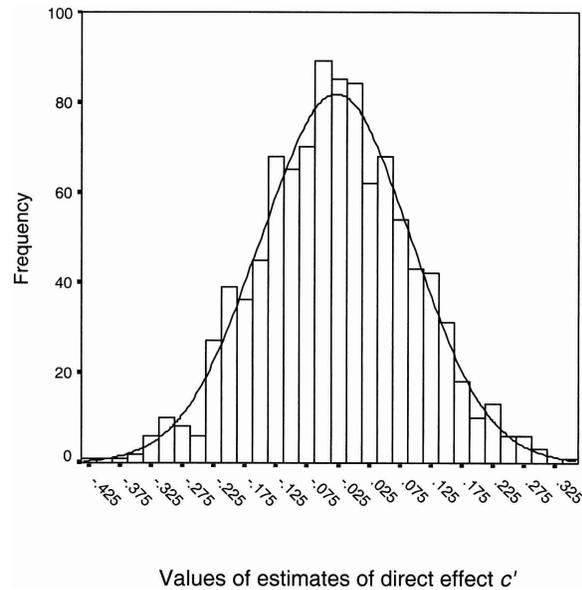
calculated 1,000 times in the bootstrap samples, the mean of the distribution was −.029, the standard deviation was .12, and the shape of the distribution is clearly symmetric and normal. In this case, it makes no difference whether one computes a normal-theory confidence interval using the bootstrap mean and standard deviation or identifies the values that select the observations in the middle 95% of the distribution. Both bootstrap methods lead to the interval (−0.27, 0.22). Although this interval is somewhat narrower than the confidence interval in the original sample, it leads to essentially the same substantive conclusion.[11] Thus, as we stated earlier, it is because the indirect effect *a* × *b* has a distribution that is skewed away

---

[10] Essentially the same distribution was obtained with 500 bootstrap samples, suggesting that the procedure has led to a stable estimate of the distribution. The percentile confidence interval in this case is virtually identical to a more sophisticated interval procedure that adjusts for possible bias (Efron & Tibshirani, 1993, pp. 178–199).

[11] The confidence interval in the original sample is somewhat wider because the estimated standard error is slightly larger than the bootstrap estimate. This difference reflects sampling variability, as the usual estimated standard error is itself a sample statistic. As the original sample size gets larger, the standard errors and confidence intervals from the two approaches are expected to converge.

from zero that the bootstrap approach has more power than the conventional approach. In contrast, the sampling distributions of a, b, and c' are symmetrical and normal, and therefore, the bootstrap test does not emerge as more powerful.

Evaluating Mediation Models

Bootstrap procedures and confidence intervals were not widely used at the time that Baron and Kenny (1986) formulated their groundbreaking guidelines for assessing mediation. In this section, we consider how these new tools might lead to a somewhat different set of guidelines. We first consider the nature of the hypothesized mediation process, whether it is proximal or distal. We argue that if the process that is to be mediated is theoretically distal, then it may not be necessary to first test the $X \rightarrow Y$ relation by using bivariate methods. We next consider the case when the bivariate relation between X and Y appears to be weak because of a suppressor process. Finally, we consider the problem of assessing whether mediation is complete or partial.

Proximal Versus Distal Mediation Processes

The clearest examples of causal processes in general, and mediated causal processes in particular, involve temporally proximal causal components. If X is a strong causal agent of Y, then one expects a change in X to lead to a new value of Y shortly thereafter. For example, an unobtrusive instance of social support to a stressed person should lead to some sign of reduced distress by the next day (Bolger et al., 2000). A cognitive prime should lead to immediate facilitation of certain memory processes. An improved lesson presentation should lead to better performance on a test the next day.

When a causal process is temporally proximal, it is likely that the effect c in Part 1 of Figure 1 will be moderate or large in magnitude. (According to Cohen, 1988, a medium effect for regression or correlation is around .30 in standardized units, and one needs a sample size of approximately 85 to detect the effect with 80% power when using the usual two-tailed significance level of .05.) As the causal process becomes more distal, the size of the effect typically gets smaller because the more distal an effect becomes, the more likely it is (a) transmitted through additional links in a causal chain, (b) affected by competing causes, and (c) affected by random factors.

The example in Table 1 illustrates the pattern of results that one expects with a mediated, distal causal process. We demonstrate that in such cases we may be able to confidently detect mediation but not the overall or total effect. In this example, the effect of X on M and of M on Y are both medium in size (0.33 and 0.28 in the sample), and the power to detect effects of these sizes with a sample size of 80 is 86% and 72%, respectively (Cohen, 1988). However, the expected correlation between X and Y, assuming total mediation, is (.33)(.28) = .09. This is a small effect and would be unlikely to be detected with a sample of only 80. According to Cohen (1988), the power available to detect a correlation of .09 with a sample of 80 is only 12%. If one wanted to have 80% power to detect a correlation of that size, one would need a sample size of 966.[12]

When a valid mediation model is available and appropriate statistical methods are used, the power to detect a distal effect is often improved over the power to test a simple bivariate association. The test of the indirect path between X and Y benefits from knowledge of the mediation process and the fact that the more proximal $X \rightarrow M$ and $M \rightarrow Y$ associations are larger than the distal $X \rightarrow Y$ association.[13] Because the test of the $X \rightarrow Y$ association may be more powerful when mediation is taken into account, it seems unwise to defer considering mediation until the bivariate association between X and Y is established. This leads us to support recommendations to set aside

---

[12] These calculations make use of the effects as estimated from the sample. As we noted earlier, the data were generated from a model that had population parameters of 0.40 and 0.30 for the effects of X on M and of M on Y, and these population values lead to an expected indirect effect of (0.4)(0.3) = 0.12. The sample size needed to detect this effect with 80% power is 542. Although this is still very large, one can see that small differences in small correlations have large effects on the required sample size.

[13] The benefits of taking the mediated path into account are somewhat complicated. When the $X \rightarrow M$ path is large (i.e., standardized effect >.7), the multicollinearity of these two variables leads to an increased standard error for $\hat{b}$, which can make the test of the mediated effect less powerful than the bivariate $X \rightarrow Y$ test. For medium effects of $X \rightarrow M$, however, the mediated effect will be more powerful. There are other examples in data analysis of prior knowledge leading to more powerful statistical models and tests. In structural equation models, for example, the standard errors of parameter estimates are typically smaller when paths that are known to have zero weight are eliminated from the model.

the first step of Baron and Kenny's (1986) classic approach (e.g., Collins et al., 1998; MacKinnon, 2000; MacKinnon et al., 2000). Especially for distal processes, for which the usual bivariate tests of association have limited power, we recommend that the mediation analysis proceed on the basis of the strength of the theoretical arguments rather than on the basis of the statistical test of X on Y.[14]

Relaxing the requirement that $X \rightarrow Y$ be statistically significant before going on to study mediation is likely to be especially important for developmental and other researchers who track long-term processes. For example, Susser, Lin, Conover, and Struening (1991) hypothesized that children who experience out-of-home placement into foster care may be at risk for increased adolescent behavior problems and that these behavior problems put them at risk for adult substance abuse and psychopathology. The effect of out-of-home placement on adolescent problems is hypothesized to be medium in size, as is the effect of adolescent problems on adult problems. In this context, the relation of out-of-home placement to adult psychopathology and substance abuse is likely to be subtle, but it is likely to be of theoretical interest as an exemplar of one of many childhood antecedents that contribute to adult functioning. We recommend that studies of such distal processes proceed without the initial bivariate test of the long-term effects and that the test of the model-based association through the indirect path be approached through the bootstrap procedure or other recommended methods (MacKinnon et al., 2002).

When the effect to be studied is proximal to its cause, however, we believe Baron and Kenny's (1986) classic recommendation that the $X \rightarrow Y$ effect be tested has conceptual usefulness. Unless there is an a priori argument that the relation of X to Y is subtle or due to a suppression process (see below), the effect size of proximal causes is likely to be medium or large, and well-designed studies will have adequate power to detect the bivariate effect. The evidence for an association between X and Y will be more convincing if both the bivariate and the indirect associations are established. Clearly, experimentalists who wish to elaborate the mechanisms of an experimental effect need to first establish that the effect exists. In sum, we agree with others that the rigid requirement of the first step of Baron and Kenny's mediation guidelines be dropped but that it be retained in specific cases in which theory suggests that the associations are large or medium in size.

Suppressor Variable Processes

In the preceding section we assumed that path c in Figure 1 will be diminished in studies of distal processes because of the dilution of one or more mediating processes. Diminution of effects over time is not the only way for a theoretically interesting association to become empirically weak in magnitude. Even proximal effects can be substantially diminished if they are suppressed by a competing process (see Cohen & Cohen, 1983, pp. 95–96; see also Tzelgov & Henik, 1991). Suppression occurs when the indirect effect $a \times b$ has the opposite sign of the direct effect c' in Figure 1.[15]

MacKinnon et al. (2000) recently provided an excellent description of suppression in the context of mediation analyses. They pointed out that Figure 1 applies to suppression effects as well as the mediation effects discussed by Baron and Kenny (1986), and following from this, they discussed Davis's (1985) conception of suppression as "inconsistent mediation" (p. 33). Because of the importance of these points, we review them in detail here.

We first illustrate suppression with an example. Consider coping behavior that is initiated because of exposure to a stressor. In this instance, X might be the occurrence of an environmental stressor (e.g., a flood), M might be the level of coping behavior (e.g., support seeking), and Y might be experienced distress. In terms of Figure 1, path a is positive (exposure to the stressor leads to the initiation of coping), but path b is negative (coping leads to reduced distress), and the indirect path through M is negative (because $a \times b < 0$). If coping behavior is held constant, the stressor is likely to lead to distress, as reflected by a value of c' that is positive. The total effect of X on Y is the sum of the indirect effect $a \times b$ and the direct effect c'. In

---

[14] Collins et al. (1998) and MacKinnon and colleagues (MacKinnon, 2000; MacKinnon et al., 2000) recommended dropping the first step of Baron and Kenny (1986) for a different reason than we do. They considered inconsistent mediating variables that may have effects that go in opposite directions, so the total effect may seem to disappear. This situation is considered briefly in the next section on suppression effects.

[15] Suppression has had a more restricted definition in educational psychology (Horst, 1941) that we are not using here. The more general meaning of suppression discussed in detail by Tzelgov and Henik (1991) and used in the present article appears to be the more widely accepted one.

a population of persons who possess good coping skills, the magnitude of c' may be similar to that of $a \times b$ but opposite in sign. In this case, the total effect (which is effect c in Part 1 of Figure 1) may be close to zero (MacKinnon et al., 2000). Clearly, the bivariate effect $X \rightarrow Y$ obscures the complexity of the causal relations between these variables.

When a suppression process is hypothesized, one may still use the mechanics of the mediation analysis that we discussed above. The indirect path $a \times b$ is meaningful and can be studied using both standard regression methods and bootstrap techniques. The direct effect c' is also meaningful but should be interpreted as a conditional effect (holding constant M; MacKinnon et al., 2000). The primary caveat is that the total effect (c in Part 1 of Figure 1) is not explained by decomposing it into two additive parts with the same sign. In fact, the suppression system can explain why a theoretically interesting relation is not strong. As we discuss in detail later, care needs to be taken when talking about total effects in the presence of suppression.

Even when a suppression process is not of initial theoretical interest, it may be observed in empirical results (see also MacKinnon et al., 2000). This is well illustrated by the numerical example shown in Table 1. The total effect of X on Y in the original sample is $c = 0.06$, which is somewhat smaller than the indirect effect of $a \times b = 0.09$. The direct effect of X on Y, holding M constant, is -0.03. In this case, the mediated effect explains more than 100% of the total effect.

This numerical example is, of course, artificial. It is a sample from a simulated population in which c' was exactly zero, and thus we know that these data should show complete mediation. Because we have unusual prior information about the population, we can be confident that the estimate of c' is negative simply because of sampling fluctuations. Both confidence intervals for c' shown in Table 1 indicate that the sampling distribution of c' ranges from around -0.3 to 0.2, given that the sample size is only 80. This is a remarkably broad range, when one considers that the indirect effect is only expected to be $(0.4 \times 0.3) = 0.12$. From this example we can see that empirical suppression will occur about 50% of the time when a model can be specified that totally accounts for a causal effect. Thus, if the population $c' = 0$, then we can expect sample estimates of c' to be negative about half of the time. As MacKinnon et al. (2000) noted, this is not true suppression, and hence we call it "empirical suppression." Our example illustrates that the degree of empirical suppression due to sampling fluctuations can be striking.

If our theory predicts that c' is near zero, then we can invoke traditional hypothesis-testing conventions to argue that these data are consistent with the null hypothesis.[16] The confidence intervals from both traditional methods and the bootstrap include zero, and a t test of the sample result $c' = -0.03$ would be far from statistically significant. In this case, we might argue that the data are consistent with total mediation. In a later section, we discuss how a bootstrap analysis might tell us how confident we can be of that claim.

In some cases, one might expect that $c' = 0$ but find that the empirical estimate of c' is negative and significantly different from zero. When this occurs, one must think carefully about the possibility of suppression effects or other alternative models rather than of the original additive mediation model. Although it is tempting to think of unexpected suppression effects as artifacts, it is our experience that they can often be informative, if not interesting.

Unexpected suppression effects sometimes occur in the context of multiwave longitudinal studies when growth processes are analyzed using simple regression models. For example, imagine a situation in which marital conflict at two points in time is used to predict marital dissatisfaction at a third time. When considered as pairs of variables, we expect that conflict (Time 1) should. be positively related to dissatisfaction (Time 3), as should conflict (Time 2), and that the two conflict measures should also be positively related. These expectations might lead us to hypothesize that the effect of conflict (Time 1) might be mediated by conflict (Time 2) and that the direct effect of conflict (Time 1) on dissatisfaction (Time 3) would be nearly zero.

In some data sets, however, suppression is found instead. When conflict is held constant at Time 2, higher levels of conflict at Time 1 seem to reduce dissatisfaction at Time 3. Mathematically speaking, this occurs when the magnitude of the overall correlation between conflict (Time 1) and dissatisfaction (Time 3) is less than one expects given the known indirect path, conflict (1) $\rightarrow$ conflict (2) $\rightarrow$ dissatis-

---

[16] If another researcher has a theory that $c' = 0.20$ or that $c' = -0.20$, then the confidence intervals also remind us that the data are consistent with that alternative theory as well.

faction (3). Psychologically speaking, such patterns of correlations can occur when the Time 1 and Time 2 measures of conflict reflect trajectory processes, whereby some couples are steadily increasing in conflict while others are reducing their conflict. The proper interpretation of such trajectory patterns has been discussed by Rogosa (1988; Rogosa, Brandt, & Zimowski, 1982). In a nutshell, when conflict at Time 2 is held constant at its mean value, then persons with unusually low Time 1 conflict will tend to have a trajectory of increasing conflict, whereas persons with unusually high Time 1 conflict will tend to have a trajectory of decreasing conflict. If the slope rather than the level is what predicts marital dissatisfaction, then the suppression pattern will be observed. The point that we make here is that some unexpected suppression results can reflect the need for an alternative theoretical model.

To recap, suppression models are closely related to mediation models, and they can be analyzed using the same set of methods. When an association is completely mediated in the population, we expect to see spurious suppression effects about half of the time because of sampling fluctuations. These effects, however, are unlikely to be statistically significant. When suppression effects are expected theoretically or when they are statistically significant, the total effect of X on Y in models involving suppression must be interpreted with caution.

Partial Mediation: Guide to Unexplained Effects

Suppression effects are observed when the direct effect c' in Figure 1 is opposite in sign to the indirect effect $a \times b$. A more common result is a value of c' that is different from zero but in the same direction as $a \times b$. As we mentioned before, this is what Baron and Kenny (1986) and Kenny et al. (1998) call "partial mediation."

Partial mediation can be observed in at least four situations. The first is the most obvious but may not be the most common. This is the situation when X has a specific and direct effect on the outcome Y in addition to its indirect effect through the mediator M. For example, suppose we were studying the effect of interpersonal conflict (X) between persons in a romantic relationship on their appraisals of intimacy (Y) the next day. The conflict may have a partially mediated effect through conscious, social–cognitive processes (M) that occur between the time of the conflict and the time of the intimacy rating, but it may also have a

direct and specific path because of affective states that were induced at the time of the conflict. In this situation, we say that the mediation model is properly specified and that the estimates of a, b, and c' are unbiased relative to population values. This situation is shown in the top panel of Figure 4.

The second situation that produces partial mediation patterns in the data is only subtly different from the first. This is the situation when there are several processes that taken together completely mediate the relation between X and Y (see Bollen, 1987; MacKinnon, 2000), but only a subset is specified explicitly in the mediation analysis. In this case, the direct effect of X on Y (c') reflects the effects that are not captured by the mediators included in the model. We illustrate this situation with an example discussed previously. Suppose X is study time, Y is achievement performance, M_1 is vocabulary acquisition, and M_2 is acquisition of technical relationships. Achievement performance is likely to be affected by both mediators. A model that omits one will have a c' path that is nonzero. In this case, the path c' can be interpreted as a kind of residual effect of X in addition to its indirect effect through M_1. We do not really expect that study time has a direct effect that is specifically comparable to vocabulary acquisition but rather that there is some other way for study time to affect performance in addition to vocabulary enrichment. Technically, the model that excludes M_2 is statistically misspecified, and the estimate of b can be biased. This is because the path from M_1 to Y is not adjusted to take into account a possibly correlated process through M_2.[17] This situation is illustrated in the middle panel of Figure 4.

The third situation that leads to the impression of partial mediation is another example of model misspecification. It sometimes happens that different mediation mechanisms apply to different persons in the population, but we are not able to anticipate this fact when designing the study. This situation has been discussed by Collins et al. (1998). Specifically, suppose that there is a group of subjects for whom M completely mediates the $X \rightarrow Y$ relation, but there is

---

[17] To simplify exposition, Figure 4 does not show that M_1 and M_2 could be correlated other than by their common cause X. In general, d_1 and d_2 could also be correlated. If M_1 and M_2 are positively correlated, such as would be expected of vocabulary and relationship acquisition, then the estimate of b (the effect of M_1 on Y) will be somewhat too large.
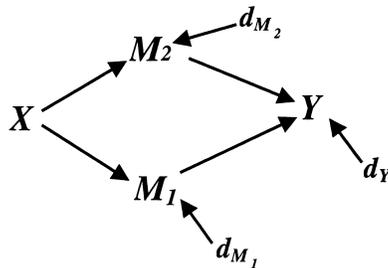
Situation One:

X **has specific and direct effect on** *Y* **in addition to an indirect effect through** *M*.
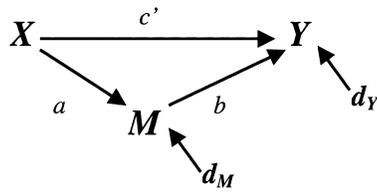
Situation Two:

X **has indirect effects on** *Y* **through both** M_1 **and** M_2, **but only** M_1 **is included in the analysis.**

Situation Three:

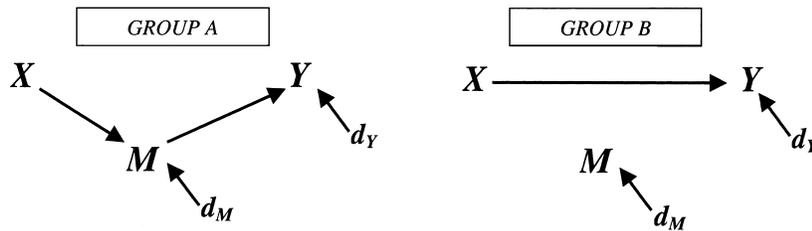**Moderated mediation: Complete mediation in Group A and no mediation in Group B.**

Figure 4. Three situations in which partial mediation may be suggested by the data (c' is reliably different from zero).

another group for which *M* does not operate. If the groups are equally represented in the population, then the analysis of the model in Figure 1 will lead to an inference that about half of the effect of *X* on *Y* is mediated through *M*. Unlike the first situation in which that ratio might apply to each and every subject in the study, this third situation is informative only as a statistical average of heterogeneous effects. The statistical model in this case is technically misspecified because it does not take into account the interaction between group and mediation process (Baron & Kenny, 1986; described as moderated mediation effects by James & Brett, 1984). This situation is shown in the bottom panel of Figure 4.

For example, Nolen-Hoeksema, Larson, and Grayson (1999) have argued that the relation between chronic strain and depressive symptoms is more likely to be mediated by rumination processes in women than in men. An analysis that pooled men and women when studying mediation of this sort would find that rumination is a partial mediator on average, even though it may explain all of the association in women and little of the association in men.

A fourth situation, not shown in Figure 4, that can

lead to partial mediation is when the variable M is measured with error. This is another example of model misspecification, but it reflects a methodological artifact rather than an opportunity to refine theory. When the mediator is measured with error, its associations with X and Y are underestimated, and hence the indirect effect would also be underestimated (Bollen, 1989).

Regardless of the source and interpretation of partial mediation effects, their presence in the data suggest that the causal mechanism is more, rather than less, complicated. With the exception of the fourth situation, these complications have the potential of enriching both theory and practice, and hence it is unfortunate that researchers often categorically dismiss partial mediation effects on the basis of null-hypothesis tests. Under this convention, authors test whether c' is statistically different from zero and, if not, declare that complete mediation has been observed. This practice ignores the fact that the test of c' often has limited statistical power and that the data are consistent with important partial mediation as well as with complete mediation.

The disadvantages of rigid adherence to null-hypothesis statistical tests have been extensively documented by Cohen (1994) and others (e.g., Carver, 1978, 1993) and are not repeated here. Instead, we simply encourage the use of the logical alternative to null-hypothesis statistical tests, the 95% confidence interval. As we discussed earlier, confidence intervals for c' can be constructed using both normal theory (Equation 1) and bootstrap distributions of $\hat{c}'$. For example, Table 1 shows numerical results that suggest that the data are consistent with parameter values of c' in the range (−0.30, 0.25). Although we note that the data are consistent with complete mediation, we should also note that the sample size of 80 does not allow us to rule out the possibility of direct effects as large as 0.25. Given that the total effect of X on Y in the population is 0.12, one can appreciate that this sample does not allow one to say with much certainty that complete mediation is established.

Quantifying Strength of Mediation

Whether or not $\hat{c}'$ is significantly different from zero, it may be of interest to represent the strength of mediation on a continuum, rather than as the partial or complete dichotomy. One approach that has been taken is to compute the ratio of the indirect effect over the total effect of $X \rightarrow Y$. We call this ratio P_M for effect proportion mediated. The notion is that the total

effect is mediated by one or more processes and that the effects of these mediation processes add up to the total effect. In the case of a single mediator such as that shown in Figure 1, P_M can be estimated using estimates of $a, b,$ and c:

$$\hat{P}_M = (\hat{a} \times \hat{b})/\hat{c}, \qquad (4)$$

where a and b are estimated from Part 2 of Figure 1 and c is estimated using Part 1. MacKinnon et al. (1995) studied $\hat{P}_M$ and derived an estimate of its standard error. In simulation studies of P_M estimates, they found that large samples (e.g., >500) were needed to obtain estimates of P_M with acceptably small standard errors.

One reason that estimates of P_M may have a large standard error is that empirical suppression effects may result in values of $\hat{c}$ that are smaller than estimates of $\hat{a} \times \hat{b}$ or even different in sign. The interpretation of P_M is clear only when the mediated and direct effects are the same sign. If suppression occurs, estimates of P_M can exceed 1.00 or even become negative.

We recommend that the effect ratio be computed only when there is no strong evidence of suppression. If slight empirical suppression is unexpectedly observed but is not statistically significant, we recommend that P_M be set at the upper bound of 1.00. The data displayed in Table 1 illustrate such an instance. The indirect effect is $\hat{a} \times \hat{b} = 0.09$, the direct effect is −0.03, and the total effect is $\hat{c} = 0.06$. However, given that c' is not significant, there is no strong evidence that suppression exists. In this case, we would compute the point estimate of the ratio as 1.00, indicating that the data set displays complete mediation.

We know that the example in Table 1 arises from a population model in which complete mediation is exhibited, but we also know from the confidence interval for c' that a sample size of 80 cannot provide assurance that partial mediation does not occur. We can use the bootstrap samples to investigate the sampling variation of the effect ratio, defined assuming no real suppression. Figure 5 shows the distribution of the ratios computed from the bootstrap samples for the data set of Table 1. The distribution is highly skewed, with more than half of the samples producing ratios of 1.00, and the other samples producing values that tail off over most of the possible range. The long tail on this distribution reminds us of MacKinnon et al.'s (1995) result that the effect ratio is a highly variable quantity to estimate. Although this sample shows complete mediation, and more than half of the boot-
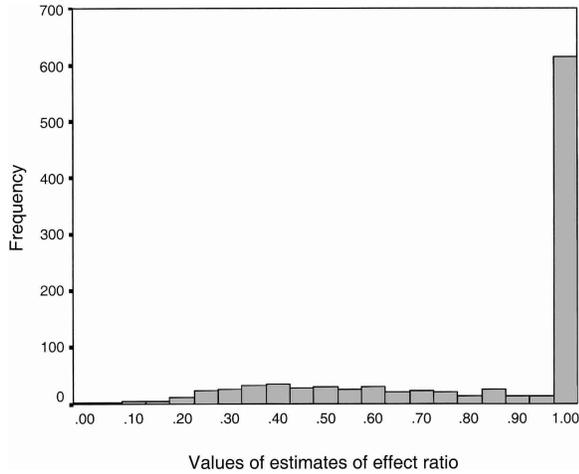
Figure 5. Bootstrap distribution of effect ratio. $SD = .28$; $M = .82$; $N = 1,000$.

strap samples are consistent with complete mediation, our 95% confidence interval indicates that the ratio could be as small as 0.22. The flattened shape of the distribution suggests that a convention of a 90% or even an 80% confidence bound for the effect ratio would be useful. The 90% lower bound in this case is 0.28, and the 80% lower bound is 0.34. This sample size of 80 is strongly consistent with complete mediation, but it is also consistent with partial mediation with as little as one third of the effect explained by the mediating process, as indicated by the lower bounds of the confidence intervals.

Illustration With Empirical Example

We illustrate these computations with an example of a published mediation study that has a slightly smaller sample size than our first example. Chen and Bargh (1997) were interested in demonstrating that subliminal primes given to a participant (the observer) could affect his or her verbal behavior toward another participant (the target) and that this hostility would in turn evoke increased hostility from the target. They studied 46 dyads, each of which consisted of one observer and one target, who interacted verbally without seeing each other. The observers were exposed subliminally to male faces, half to an African American and half to a Caucasian American. The verbal behavior was evaluated blindly by independent raters for hostility on a 7-point scale ranging from 1 (*super nice*) to 7 (*high levels of outward hostility*). Chen and Bargh expected the subliminal priming of the observer to affect the behavior of the target in the following way: The African American prime would make the observer more hostile, and the observer's verbal hostility would make the target more verbally hostile. Thus, the observer's hostile behavior would mediate the relation between the prime and the target's hostile behavior. This is what they found: The observers who were exposed to the African American face had higher levels of hostility, and this appeared to mediate a relationship between the prime and the hostility of the target.

The results of the mediation analysis are shown in Table 2. Path *a* represents the effect of the prime on the observer's behavior. The observers who were exposed to the African American prime tended to have hostility scores 0.34 units (on a 7-point scale) higher than those who were exposed to the Caucasian prime. Both the normal-theory confidence intervals and the bootstrap percentile interval give the sense that this effect is reliable. Path *b* represents the effect of ob-

Table 2

Illustration of Standard and Bootstrap Mediation Methods Using Chen and Bargh's (1997) Data on the Effect of Observer's Subliminal Priming (X) on Target's Behavior (Y) as Mediated Through Observer's Behavior (M)

| | | | 95% CI | | | Bootstrap | |
| | | | Standard normal | Bootstrap percentile | Bias-corrected | | |
Effect	Estimate	SE				M	SD
a	0.342	0.159	(0.02, 0.66)	(0.04, 0.67)	(0.03, 0.67)	0.347	0.155
b	0.429	0.172	(0.08, 0.77)	(0.03, 0.82)	(0.04, 0.82)	0.424	0.191
c'	0.158	0.190	(−0.22, 0.54)	(−0.23, 0.51)	(−0.20, 0.53)	0.148	0.188
a × b	0.147	0.089	(−0.03, 0.32)	(−0.001, 0.49)	(0.004, 0.52)	0.160	0.121
c	0.304	0.192	(−0.08, 0.69)				

Note. Estimates are unstandardized. The prime is coded as a binary (0,1) variable, and perceiver and target hostility are coded on a 7-point (1–7) scale. Unlike Chen and Bargh, we did not adjust for gender because it had no relation to any of the other variables. Normal 95% CIs are computed using estimate ±2 × SE. Percentile 95% CIs for bootstrap distributions are defined using the values that mark the upper and lower 2.5% of each distribution. CI = confidence interval.

server's behavior on target's behavior. In Chen and Bargh (1997) data, on average, one unit change in the hostility of the observer was associated with a 0.43-unit higher hostility score in the target. Again, the normal-theory confidence intervals and the bootstrap percentile interval give similar results, although in this case the intervals based on the bootstrap tend to be somewhat wider than those based on the ordinary least squares analysis. Path c' represents the direct effect of the prime on the target's behavior after controlling for the observer's behavior, and path c is the total effect of the prime on the target's behavior without adjusting for the observer's behavior. The former is about half the size of the latter, and all the forms of the confidence interval range from −0.20 to 0.50. The data are consistent with no direct effect, but they are also consistent with relatively large direct effects.

Figure 6 shows the bootstrap distribution of the indirect effect $a \times b$. The average indirect effect was about 0.16, and both the normal-theory and the percentile intervals include zero. Chen and Bargh (1997) noted that a one-tailed, normal-theory significance test (corresponding to the 90% confidence bound) leads to the conclusion that the indirect effect is nonzero and that observation holds for the percentile results as well. Although the 95% interval includes zero,
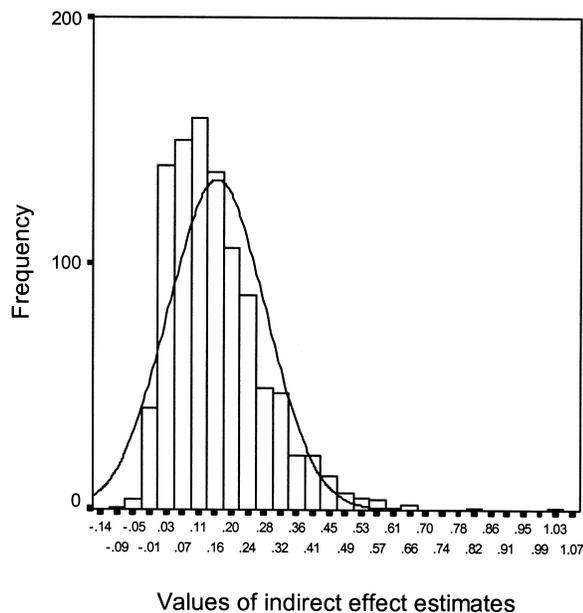
the lower bound is much closer to zero than the normal-theory lower bound. Figure 6 shows that the distribution of $a \times b$ is distinctly skewed in this example and that an assumption of normality will lead to inaccurate probability statements.

The last column of Table 2 gives a slightly different result from the intervals based on the standard normal and percentile methods. This is a bias-corrected bootstrap confidence interval that produces more accurate confidence intervals with small samples (see Efron & Tibshirani, 1993, pp. 178–188).[18] In the first example with $N = 80$, the refinement did not make any difference, but in this case, the bias-corrected interval for the indirect effect no longer includes zero. With the skewed distributions of the indirect effect estimates, the application of the bias-corrected interval method tends to improve the power of the test of the indirect effect.

Whereas the conventional analytic methods available to them led Chen and Bargh (1997) to conclude that the effect of priming on target behavior was completely mediated by observer behavior, the analysis with bootstrap methods and confidence intervals suggests that these data are consistent with partial mediation as well as with complete mediation. Figure 7 shows the distribution of the ratio of mediated effects to total effects, assuming no genuine suppression. It is clear that the sample size does not allow much closure to be reached regarding how much of the effect of priming on target behavior is mediated by observer behavior. Both the mean and the median of this distribution are around 0.50, and the bootstrap 95% confidence interval ranges from −0.02 to 1.00. The 90% interval ranges from 0.05 to 1.00. The best guess at this point is that about half the effect of the prime on target hostility is mediated by observer hostility, as detected by the judges rating the nonverbal behavior, but complete mediation may also be consistent with

Figure 6. Bootstrap distribution of indirect effect ($\hat{a} \times \hat{b}$), Chen and Bargh's (1997) example. $SD = .12$; $M = .16$; $N = 1,000$.

[18] The computational details of the bias-correction adjustment are beyond the scope of this article, but, in brief, takes into account skewness of the bootstrap distribution and the estimated change in the standard error of the parameter as a function of the presumed parameter value. The bias-correction adjustment is not a closed-form equation but rather an algorithm that makes use of resampling. Efron and Tibshirani (1993) provided the details as well as evidence of the improved accuracy of the adjustment. The algorithm is implemented in the AMOS statistical software system (Arbuckle, 1999).
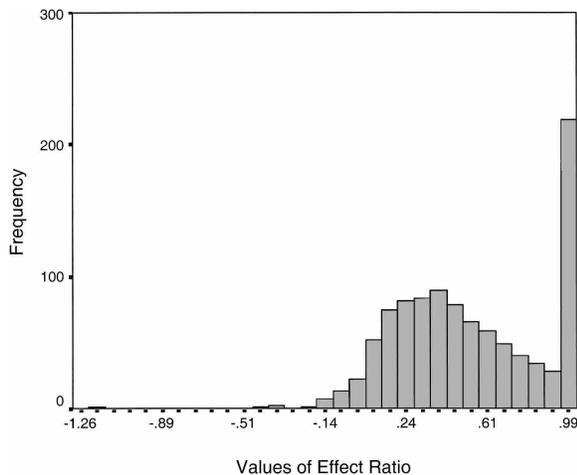
Figure 7. Bootstrap distribution of effect ratio, Chen and Bargh's (1997) example. $SD = .33$; $M = .54$; $N = 1,000$.

the data. The bootstrap confidence intervals are consistent with MacKinnon et al.'s (1995) warning that P_M has extensive sampling variation when samples sizes are small, even if the distribution is adjusted so that $\hat{P}_M$ does not exceed 1.00.

Practical Computation of Bootstrap Distributions

Bootstrap samples are conceptually easy to construct, as they are simply samples with replacement from an existing data set. However, the practical computation of large numbers of bootstrap samples, and the repeated application of estimation methods to these samples, requires special computer programs. Lockwood and MacKinnon (1998) have shown how these programs can be written for analyses of mediation models using the SAS statistical software (SAS Institute, 1989), and Efron and Tibshirani (1993) recommended S-PLUS (StatSci, 1999).

For the substantive researcher who is interested in mediation models, we recommend the use of the structural equation packages EQS (Bentler, 1997) and AMOS (Arbuckle, 1999), which are commonly available in academic settings. These packages now include bootstrap options that make the formation of bootstrap samples and the estimation of results easy to compute. Annotated examples of programs that created the results from Table 1 (EQS) and Table 2 (AMOS) are shown in the appendixes. For both illustrations, 1,000 bootstrap samples were created. In the

EQS example, we save the 1,000 sets of estimates, which can be summarized using any statistical software. Examples of SPSS (SPSS Inc., 1999) syntax for reading the bootstrap estimates, creating the effect-ratio estimates, and plotting results are also provided in Appendix B.[19] If one is not computing the effect ratio, then AMOS may be more convenient at this time. Results are available in one step, and one can request bias-corrected confidence intervals.

New Recommendations for Analysis of Mediation Processes

Figure 8 is a schematic representation of our guidelines for assessing mediation. It is a revision of Baron and Kenny's (1986) and Kenny et al.'s (1998) guidelines that explicitly takes into account the strength of the effect to be mediated (distal vs. proximal) and the possibility that suppression might be involved. Although not explicit in Figure 8, for Step 4 we suggest that researchers make use of new statistical tools for studying relations in mediation/suppression models, including bootstrap methods for studying variation of parameter estimates and confidence intervals for recognizing that small samples yield shades of gray rather than black-and-white results.

The first step is to think conceptually about the nature and size of the effect to be mediated. If the effect can be experimentally induced, it is usually proximal, and one should be able to demonstrate it by showing an association of the causal variable X with the outcome Y. We put this step in a dashed box in Figure 8 to indicate that the step is not rigidly required. If, for example, there is good reason to believe the statistical power to detect $X \rightarrow Y$ is inadequate, one may proceed to Step 2. In this case, one would expect that the confidence interval around the c effect includes effect sizes that are predicted by the theory being examined, even if the confidence interval also includes zero. Our suggestion that Step 1 not be rigidly held in investigations of proximal effects should not take away from the fact that the most convincing evidence of a causal effect will always follow from demonstrating that an experimentally manipulated X has a reliable association with the other variables in Figure 1.

---

[19] Readers may download these programs and the data set used in Table 1 from the following Web site: www.psych.nyu.edu/couples/PM2002.

```
            ┌─────────────────────────────┐
            │  Specify theoretical expectations
            │  of intervening variable process.
            └─────────────────────────────┘
```

Figure flow (text content of boxes):

- **Specify theoretical expectations of intervening variable process.**

- (oval) **Mediation of Proximal Effect?**
- (oval) **Mediation of Distal Effect?**
- (oval) **Suppression?**

- (dashed box) 1. Estimate/test bivariate $X \rightarrow Y$ path (c).

- (box) 2. Estimate/test bivariate $X \rightarrow M$ path (a).
  3. Estimate/test $M \rightarrow Y$ path with $X \rightarrow Y$ held constant (b).
  4. Estimate/test indirect $X \rightarrow M \rightarrow Y$ path ($a \times b$).
  5. Estimate/test $X \rightarrow Y$ path with $M \rightarrow Y$ held constant (c').

- (box) Same as steps 2 to 5.

- (box) Is there **strong** evidence that direct path, c', is opposite sign as indirect path $a \times b$?

- (box) Is there **any** evidence of suppression?

- **NO** / **YES** / **YES**

- (box) 6. Assess strength of mediation using effect ratio.
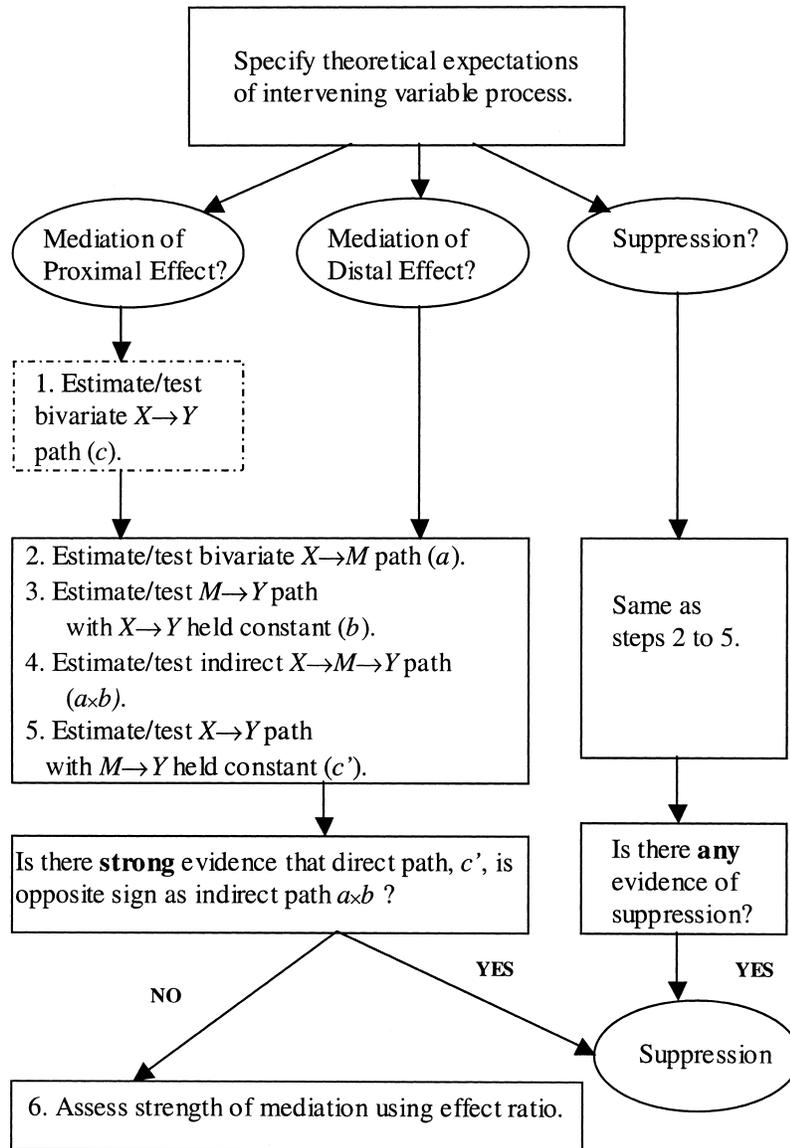
- (oval) Suppression

Figure 8. Schematic decision tree for mediation analysis. The dashed box indicates a step that is recommended but not required. Solid boxes indicate steps that are strongly recommended.

If the $X \rightarrow Y$ effect is more distal and the magnitude of the expected effect is small, we recommend proceeding directly to the second step of Baron and Kenny's (1986) procedure, rather than risking a Type II error for the entire mediation system. If the association between X and Y can be explained with a model that involves suppression, then one also proceeds to the second step of the causal process analysis. As we noted above, we are not the first to recommend that Step 1 of Baron and Kenny's analysis be suspended.[20]

In addition to the formal statistical tests of effects, Figure 8 reminds researchers to attend to estimation, both point estimates and interval estimates. The point estimates of a, b, c', and c give important information

_____

[20] The reviewers of this article had mixed opinions about whether any form of Step 1 should be retained. Two believed it should be dropped completely. Another argued for retaining the step because it provides protection against alternative causal models, whereby the associations of X and M and of M and Y are spurious. We believe our suggestion

about the potency of the mediation,[21] and the confidence intervals help to gain perspective on how definitive are the results. The confidence intervals on these regression coefficients can be computed using usual normal-theory methods when the variables M and Y are quantitative, or they can be computed more generally using the bootstrap methods described above.

The product of the point estimates of a and b describes the indirect effect of X on Y through mediator M, and the confidence interval for this quantity should be computed using percentile methods applied to the bootstrap distribution. In many cases, the lower bound of this confidence bound will exclude zero even if the total effect of X on Y is not significantly different from zero. The increased power of this test derives from the assumption that the mediation model is correct.

If no suppression exists theoretically or empirically, then Figure 8 recommends that the strength of the mediation effect be estimated using the ratio of the indirect effect to the total effect. This quantity is subject to a large amount of sampling variation, however, so we also recommend that confidence intervals be computed using the percentile bootstrap methods. In most cases, this analysis will lead to conclusions that have shades of gray rather than being black or white.

Concluding Comments

In this article, we have focused on three-variable mediation models, in which the relation between a pair of variables is mediated by a third variable. In doing so, we made a number of assumptions that deserve reiteration. We assumed that the independent variable, X, and the mediator variable, M, were measured without error and that the hypothetical causal sequence $X \rightarrow M \rightarrow Y$ was plausible, if not compelling. We also assumed that the relations among the variables could be represented using linear models. Even when all of these assumptions are met, we acknowledge that statistical mediation analyses based on nonexperimental data provide suggestive rather than definitive evidence regarding causal processes. We did not consider studies that experimentally manipulate M, but such designs are worth pursuing if statistical evidence of mediation is observed.

that Step 1 be recommended for proximal effects, but not be essential for distal effects, will still encourage investigators of proximal effects to use study designs that have sufficient statistical power to detect such effects.

The bootstrap methods that we recommend for testing indirect effects complement the methods recently studied by MacKinnon et al. (2002) in extensive simulations. Although they did not include bootstrap-based tests in their simulations, they showed that the usual tests based on large-sample standard errors are too conservative. They recommended, with caveats, some methods that use information about the exact distribution of the products of two normal variables (Meeker, Cornwell, & Aroian, 1981). The bootstrap method can be viewed as an approximation of this exact distribution.

In principle, the bootstrap methods can be generalized to mediation analyses that make use of structural equation methods with measurement models for the independent, mediator, and outcome variables. Measurement models can be used to adjust for measurement error, and given that measurement error in the

_____

[21] A reviewer questioned whether it is wise to estimate c' routinely. The reviewer strongly endorsed Rozeboom's (1956) formulation of mediation, whereby the effect of M on Y would be studied without adjusting for X in instances when the assumption of complete mediation (i.e., $X \rightarrow M \rightarrow Y$) is warranted. From this perspective, Step 3 would be modified if there is evidence that c' (tested in Step 5) is zero. We agree that if the complete mediation model can be established definitively then the test of effect b will be more powerful in a structural model that does not include c'. As Kenny et al. (1998) have discussed, this increase in power will be especially apparent when the effect of X on M is large. However, we resist endorsing a general recommendation that $M \rightarrow Y$ be tested without adjusting for X, because this test can be seriously biased whenever partial mediation is plausible, whether or not c' is statistically significant. For example, suppose that in the population both X and M have medium effects of .30 on Y, but X has a strong effect of .90 on M. The partial mediation model would reveal that slightly less than half of the variation goes through the indirect effect. Assuming complete mediation would be clearly incorrect, and the biased estimate of the $M \rightarrow Y$ effect would be around .57 rather than .30. However, there would be insufficient power (< 90%) to detect the $X \rightarrow Y$ direct effect with these effect sizes unless sample size exceeded 400. In this example, investigators with samples in the 100–200 range would be at risk of committing a Type II error if they tested the $X \rightarrow Y$ direct effect and failing to find significance, dropped the path from the model. Their conclusion of complete mediation would be incorrect, and their estimate of the $M \rightarrow Y$ path would be strikingly biased. The guidelines in Figure 8 guard against this error.

proposed mediator can lead to spurious inferences of partial mediation, this extension is of great practical importance. Although these more complex models may raise additional issues about model fit and convergence, the software discussed in Appendix B can be readily used to estimate them.

The methods we discuss also generalize to other structural models besides those shown in Figure 1. For example, if complete mediation is assumed rather than tested, one would estimate a model that constrains c' in Figure 1 to be zero. Bootstrap methods can still be used to study the distribution of $\hat{a} \times \hat{b}$ from this constrained model. The same methods can also be used with models in which there are multiple mediators. For instance, placement of children in foster care systems may lead to patterns of adolescent behavior problems through multiple mediating paths. The calculation of the size of the indirect effect in these cases is more complicated, and it involves the sum of alternative indirect pathways (see Bollen, 1989, pp. 376–389). Nonetheless, the bootstrap distributions of the indirect pathways can be calculated just as before.[22] The issue of suppression may be more complicated in a multiple-mediator model, because each mediated effect is adjusted for the other effects, and these mediating processes may be correlated. Care should be taken when interpreting the ratio of the total effect that is mediated if suppression effects are present.

The use of the bootstrap to estimate and to test direct and indirect effects can help with mediation problems in which the mediator and outcome variables are not normally distributed. The two examples we studied satisfied the usual assumptions that error distributions are normal, and we noted that the normal-theory results and the bootstrap gave virtually the same results for tests of direct effects. If we were studying mediator variables that were binary or ordinal, then the normal-theory and bootstrap results might begin to diverge. In these cases, the bootstrap results are likely to yield more accurate probability estimates. In the future, we expect that bootstrap methods will become widely available in general-purpose statistical software as well as in software for structural equation models.

The wide applicability of bootstrap methods to models involving variables that are not normally distributed does not mitigate the need to justify the specification of the mediation model. We have assumed simple linear regression models in our discussion, but more complicated nonlinear alternatives should be considered for binary variables (Collins et al., 1998; Kraemer, Stice, Kazdin, Offord, & Kupfer, 2001). Bootstrap methods can be used with nonlinear models as well as the linear models we have considered in our illustrations.

Because bootstrap methods have many desirable features, we have argued that they are particularly useful in studying indirect effects in mediation models. The improvement of bootstrap-based inference over normal-theory approximations is most obvious when one is interested in rejecting the null hypothesis of no indirect effect. Substantive researchers are not generally aware that mediated associations can be shown to be reliable when the simple bivariate correlation is not significant. One reason this fact might not be widely known is that standard tests, such as Sobel's (1982) test, do not fully capitalize on the power afforded by the intervening variable. By taking into account the skewed shape of the sampling distributions of indirect effects, investigators' use of bootstrap-based tests is better able to advance a goal of psychological research, testing theories of mediating processes.

---

[22] All structural equation programs calculate indirect effects and large-sample standard errors, and AMOS 4.0 reports the bootstrap distributions of the indirect effects. To study the bootstrap distribution of indirect effects from multiple mediator models using EQS, one must compute the indirect effect estimates in a separate step, such as the SPSS step shown in Appendix B.

References

Arbuckle, J. L. (1999). AMOS 4 [Computer software]. Chicago: Smallwaters Corp.

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. (1997). EQS for Windows (Version 5.6) [Computer software]. Los Angeles: Multivariate Software, Inc.

Bolger, N., Zuckerman, A., & Kessler, R. C. (2000). Invisible support and adjustment to stress. *Journal of Personality and Social Psychology, 79,* 953–961.

Bollen, K. A. (1987). Total direct and indirect effects in structural equation models. In C. C. Clogg (Ed.), *Sociological methodology* (pp. 37–69). Washington, DC: American Sociological Association.

Bollen, K. A. (1989). *Structural equations with latent variables.* New York: Wiley.

Bollen, K. A., & Stine, R. (1990). Direct and indirect effects: Classical and bootstrap estimates of variability. *Sociological Methodology, 20,* 115–140.

Carver, R. P. (1978). The case against statistical significance testing. *Harvard Educational Review, 48,* 378–399.

Carver, R. P. (1993). The case against statistical significance testing, revisited. *Journal of Experimental Education, 61,* 287–292.

Chan, W., Ho, R. M., Leung, K., Chan, D. K.-S., & Yung, Y.-F. (1999). An alternative method for evaluating congruence coefficients with Procrustes rotation: A bootstrap procedure. *Psychological Methods, 4,* 378–402.

Chen, M., & Bargh, J. A. (1997). Nonconscious behavioral confirmation processes: The self-fulfilling consequences of automatic stereotype activation. *Journal of Experimental Social Psychology, 33,* 541–560.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences.* Hillsdale, NJ: Erlbaum.

Cohen, J. (1994). The earth is round ($p < .05$). *American Psychologist, 49,* 997–1003.

Cohen, J., & Cohen, P. (1983). *Applied multiple regression/ correlation analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Collins, L. M., Graham, J. W., & Flaherty, B. P. (1998). An alternative framework for defining mediation. *Multivariate Behavioral Research, 33,* 295–312.

Conger, R. D., Elder, G. H., Jr., Lorenz, F. O., Conger, K. J., Simons, R. L., Whitbeck, L. B., et al. (1990). Linking economic hardship to marital quality and instability. *Journal of Marriage and the Family, 52,* 643–656.

Davis, J. A. (1985). *The logic of causal order.* Thousand Oaks, CA: Sage.

Draper, N. R., & Smith, H. (1998). *Applied regression analysis* (3rd ed.). New York: Wiley.

Eagly, A. H., & Chaiken, S. (1993). *The psychology of attitudes.* Fort Worth, TX: Harcourt Brace Jovanovich.

Efron, B. (1988). Bootstrap confidence intervals: Good or bad. *Psychological Bulletin, 104,* 293–296.

Efron, B., & Tibshirani, R. (1993). *An introduction to the bootstrap.* New York: Chapman & Hall/CRC.

Holland, P. (1986). Statistics and causal inference (with discussion). *Journal of the American Statistical Association, 81,* 945–970.

Horst, P. (1941). The role of predictor variables which are independent of the criterion. *Social Science Research Council Bulletin, 48,* 431–436.

James, L. R., & Brett, J. M. (1984). Mediators, moderators, and tests for mediation. *Journal of Applied Psychology, 69,* 307–321.

Jöreskog, K. G., & Sörbom, D. (1993). LISREL 8 (Version 8.12) [Computer software]. Chicago: Scientific Software.

Judd, C. M., & Kenny, D. A. (1981). Process analysis: Estimating mediation in treatment evaluations. *Evaluation Review, 5,* 602–619.

Kenny, D. A., Kashy, D. A., & Bolger, N. (1998). Data analysis in social psychology. In D. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *Handbook of social psychology* (4th ed., Vol. 1, pp. 233–265). New York: McGraw-Hill.

Kraemer, H. C., Stice, E., Kazdin, A., Offord, D., & Kupfer, D. (2001). How do risk factors work together? Mediators, moderators, and independent, overlapping, and proxy risk factors. *American Journal of Psychiatry, 158,* 848–856.

Lee, W.-C., & Rodgers, J. L. (1998). Bootstrapping correlation coefficients using univariate and bivariate sampling. *Psychological Methods, 3,* 91–103.

Link, B. G., & Shrout, P. E. (1992). Spurious associations in longitudinal research. *Research in Community and Mental Health, 7,* 301–321.

Lockwood, C. M., & MacKinnon, D. P. (1998, March). *Bootstrapping the standard error of the mediated effect.* Paper presented at the 23rd Annual Meeting of SAS Users Group International, Cary, NC.

Maccorquodale, K., & Meehl, P. E. (1948). On a distinction between hypothetical constructs and intervening variables. *Psychological Review, 55,* 95–107.

MacKinnon, D. P. (2000). Contrasts in multiple mediator models. In J. S. Rose, L. Chassin, C. C. Presson, & S. J. Sherman (Eds.), *Multivariate applications in substance use research: New methods for new questions* (pp. 141–160). Mahwah, NJ: Erlbaum.

MacKinnon, D. P., & Dwyer, J. H. (1993). Estimating mediating effects in prevention studies. *Evaluation Review, 17,* 144–158.

MacKinnon, D. P., Krull, J. L., & Lockwood, C. M. (2000). Equivalence of the mediation, confounding and suppression effects. *Prevention Science, 1,* 173–181.

MacKinnon, D. P., & Lockwood, C. M. (2001). *Distribution of products tests for the mediated effect: Power and Type I error rates.* Unpublished manuscript.

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., Warsi, G., & Dwyer, J. H. (1995). A simulation study of mediated effect measures. *Multivariate Behavioral Research, 30,* 41–62.

Marks, M. A., Zaccaro, S. J., & Mathieu, J. E. (2000). Performance implications of leader briefings and team-interaction training for team adaptation to novel environments. *Journal of Applied Psychology, 85,* 971–986.

Mathieu, J. E., Heffner, T. S., Goodwin, G. F., Salas, E., &

Cannon-Bowers, J. A. (2000). The influence of shared mental models on team process and performance. *Journal of Applied Psychology, 85,* 273–283.

Meeker, W. Q., Cornwell, L. W., & Aroian, L. A. (1981). The product of two normally distributed random variables [Monograph]. *Selected Tables in Mathematical Statistics, 7,* 1–256.

Microsoft Inc. (1999). Excel 2000 [Computer software]. Redmond, WA: Author.

Nolen-Hoeksema, S., & Jackson, B. (2001). Mediators of the gender difference in rumination. *Psychology of Women Quarterly, 25,* 37–47.

Nolen-Hoeksema, S., Larson, J., & Grayson, C. (1999). Explaining the gender difference in depressive symptoms. *Journal of Personality and Social Psychology, 77,* 1061–1072.

Pearl, J. (2000). *Causality: Models, reasoning and inference.* Cambridge, England: Cambridge University Press.

Polansky, A. M. (1999). Upper bounds on the true coverage of bootstrap percentile type confidence intervals. *American Statistician, 53,* 362–369.

Rogosa, D. R. (1988). Myths about longitudinal research. In K. W. Schaie, R. T. Campbell, W. Meredith, & S. C. Rawlings (Eds.), *Methodological issues in aging research* (pp. 171–210). New York: Springer Publishing Company.

Rogosa, D. R., Brandt, D., & Zimowski, M. (1982). A growth curve approach to the measurement of change. *Psychological Bulletin, 92,* 726–748.

Rozeboom, W. W. (1956). Mediation variables in scientific theory. *Psychological Review, 63,* 249–264.

SAS Institute. (1989). SAS (Version 6.12) [Computer software]. Cary, NC: Author.

Sobel, M. E. (1982). Asymptotic confidence intervals for indirect effects in structural equation models. In S. Leinhardt (Ed.), *Sociological methodology* (pp. 290–312). San Francisco: Jossey-Bass.

Sobel, M. E. (1986). Some new results on indirect effects and their standard errors in covariance structure models. In N. Tuma (Ed.), *Sociological methodology* (pp. 159–186). Washington, DC: American Sociological Association.

SPSS Inc. (1999). SPSS (Version 9) [Computer software]. Chicago: Author.

StatSci. (1999). S-PLUS [Computer software]. Seattle, WA: Author.

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.

Susser, E. S., Lin, S. P., Conover, S. A., & Struening, E. L. (1991). Childhood antecedents of homelessness in psychiatric patients. *American Journal of Psychiatry, 148,* 1026–1030.

Tzelgov, J., & Henik, A. (1991). Suppression situations in psychological research: Definitions, implications, and applications. *Psychological Bulletin, 109,* 524–536.

Appendix A

Notes on the Bootstrap Method

Efron and Tibshirani (1993) described the bootstrap as "a computer-based method for assigning measures of accuracy to statistical estimates" (p. 10). The usual measure of accuracy for statistical quantities such as means, regression coefficients, and correlations is the standard error. A standard error is the expected standard deviation of the estimate if the estimation study had been repeated a large number of times. Usually a standard error is estimated from equations that are derived from a series of assumptions and mathematical operations. The bootstrap can be used to estimate the standard error using an empirical approach, rather than an explicit application of a formula.

The bootstrap method involves having a computer program generate a series of data sets that are designed to resemble the ones that would be observed if the estimation study were repeated many times. Each bootstrap data set is obtained by sampling (with replacement) from the original data. Call these data sets *bootstrap samples.*

For example, suppose $N = 8$ and the following heights in inches are observed: original sample: (68, 69, 69, 70, 71, 72, 72, 74). We use this small example to make the steps involved in using bootstrap methods very clear. The statistical validity of the bootstrap method is best when the sample size is 20 or more (Polansky, 1999).

Here are 5 bootstrap samples that were created by randomly sampling with replacement from the original sample:

Bootstrap Sample 1: (69, 69, 71, 72, 72, 73, 73, 74)
Bootstrap Sample 2: (68, 68, 69, 69, 70, 70, 71, 72)
Bootstrap Sample 3: (68, 68, 68, 69, 71, 72, 73, 74)
Bootstrap Sample 4: (69, 69, 70, 72, 72, 72, 73, 73)
Bootstrap Sample 5: (69, 70, 71, 72, 72, 72, 73, 74).

Note that the original sample has two instances each of 69- and 72-in. subjects, and the values are more likely than any other to occur more than once in the bootstrap samples. However, this tendency is not a certainty. In Sample 3, by chance the value 68 was chosen three times, and the values

69 and 72 were each chosen once. Note also that the original sample does not contain someone 73 in. tall, and thus none of the bootstrap samples have a person of that height.

The mean of the original sample is 70.63, and the estimated standard error of the mean is 0.706. The means of the bootstrap samples are 71.63, 69.63, 70.38, 71.25, and 71.63. The standard deviation of these five values is 0.88, and this is the bootstrap-estimated standard error, although we would never use a bootstrap estimate based on only 5 bootstrap samples.

The usual estimate of the confidence interval for the mean is $70.63 \pm (0.71) \times (2.37)$, leading to the bounds (68.96, 72.29), where 2.37 is the *t* value corresponding to a tail area 0.975 and 7 degrees of freedom. As we expect, all five bootstrap-sample means fall within this range. If we had constructed 100 bootstrap samples, however, we would have expected about five sample means to be either less than 68.98 or greater than 72.29, because it is well-known that the sample mean is approximately normally distributed.

If we had created 1,000 bootstrap samples, then we could have inferred the confidence region on the mean without calculating the standard error and assuming a normal distribution for the estimate. A percentile estimate of the 95% confidence interval is computed by ordering the 1,000 bootstrap-sample means from lowest to highest and marking the 25th out of 1,000 as the lower bound and the 975th out of 1,000 as the upper bound. Formally we call the *i*th-ordered estimate $\theta^{(i)}$, and we choose the confidence bounds to be $\theta^{(\alpha)}$ and $\theta^{(1-\alpha)}$, where α is chosen so that total coverage is $1 - 2\alpha$. For the 95% confidence interval, α is .025.

Although the percentile estimates are easy to calculate, it is known that they tend to be too narrow. Efron and Tibshirani (1993, pp. 184–186) described bias-corrected intervals that use bounds $(\theta^{(\alpha 1)}, \theta^{(\alpha 2)})$ instead of the percentile bounds $(\theta^{(\alpha)}, \theta^{(1-\alpha)})$. The values $\alpha 1$ and $\alpha 2$ are defined in a way that takes into account asymmetry in the distribution of bootstrap estimates. When the distribution of the bootstrap estimates is normal, the bias-corrected interval and the percentile intervals are virtually the same.

(Appendixes continue)

Appendix B

Using EQS and AMOS Software to Implement Bootstrap Analyses

To use EQS to carry out the bootstrap analysis, first open the data set in an EQS window, and save it with the *.ESS extension. Next, run PROGRAM 1 in EQS, such as the one shown below. This program saves 1,000 sets of parameter estimates from the bootstrap samples in the new file called "M3BOOT.txt". Finally, the new file is read into an SPSS program (PROGRAM 2), which computes the indirect effect and effect ratio and allows these to be plotted.

Program 1 (EQS)
```
/TITLE
Bootstrapped analysis of simulated data, OLS
! Comments can be inserted following exclamation points
 and ending in semicolons;
/SPECIFICATIONS
 DATA = 'MEDSIM1.ESS';
 VARIABLES = 3; CASES = 80;
 METHODS = ls; !Least squares converges more regularly than ML;
 MATRIX = RAW;
/EQUATIONS
V2 = + .331*V1 + E2;
V3 = + -.03*V1 + .275*V2 + E3;
!The above equations specify starting values from the full sample.
 These speed up the processing time;
/VARIANCES
V1 = .68*;
E2 = 1.1*;
E3 = .93*;
/COVARIANCES
/PRINT
 effect = yes;
 digit = 3;
 linesize = 80;
/SIMULATION
 SEED = 6062000;
 REPLICATIONS = 1000;
 BOOTSTRAP = 80;
! The 4 lines above control the bootstrap. Seed is an arbitrary number that starts the
  random number generation. Replications are how many bootstrap samples you want, while
  Bootstrap = specifies the size of each bootstrap sample;
/OUTPUT ;
 PARM; SE;
 DATA = 'M3BOOT.TXT';
/END
```

Program 2 (SPSS)

```
COMMENT this program reads as input bootstrapped distributions from EQS from ml
estimation of complete mediation simulation example.
COMMENT The estimated coefficients are called Ba, Bb and bc.
data list file = 'C:\ M3boot.TXT' fixed records = 5
 /1 cndcode 6-9(F) convge 10-13(F)
 /2 VarV1 1-16(F) VarE2 17-32(F) VarE3 33-48(F)
  Ba 49-64(F) bc 65-80(F) Bb 81-96(F).
compute indir = ba*bb.
compute ratio = indir/(indir+bc).
```

```
COMMENT assuming no true suppression, we define sampling suppression to be instances of
complete mediation.
if (bcprime le 0) ratio = 1.0.
execute.
COMMENT sometimes EQS fails to converge in a bootstrap sample—such instances are set
aside by the following line.
Select if (cndcode le 0).
DESCRIPTIVES
 VARIABLES = ba bb bc indir
 /STATISTICS = MEAN STDDEV VARIANCE MIN MAX KURTOSIS SKEWNESS .
GRAPH /HISTOGRAM(NORMAL) = indir .
SORT CASES BY indir (A) .
FREQUENCIES
 VARIABLES = indir /FORMAT = NOTABLE
 /NTILES = 4 /PERCENTILES = 2.5 5 95 97.5
 /STATISTICS = ALL /ORDER ANALYSIS .
GRAPH /HISTOGRAM = ratio .
SORT CASES BY ratio (A) .
FREQUENCIES
 VARIABLES = ratio /FORMAT = NOTABLE
 /NTILES = 4 /PERCENTILES = 2.5 5 95 97.5
 /STATISTICS = ALL /ORDER ANALYSIS .
```

To use AMOS 4.0, first open the data set in an SPSS window. Choose AMOS from the Analyze menu, and draw the desired path model. Select "Analysis properties" from the View/Set menu. Find the "Output" tab, and click on the box for "Indirect, direct and total effects." Find the "Bootstrap" tab, and click on the box for "Perform bootstrap." Enter the number of replications (1,000 or more for confidence intervals). Click on the box requesting confidence intervals. AMOS will provide information about the bootstrap distribution for all the effects.

Shrout, PE & Bolger, N (2002)
Mediation Analysis and Bootstrap Methods
Appendix A

Notes on the Bootstrap Method
Corrected March, 2003

Efron and Tibshirani (1993) describe the bootstrap as "a computer-based method for assigning measures of accuracy to statistical estimates". The usual measure of accuracy for statistical quantities such as means, regression coefficients and correlations is the standard error. A standard error is the expected standard deviation of the estimate if the estimation study had been repeated a large number of times. Usually a standard error is estimated from equations that are derived from a series of assumptions and mathematical operations. The bootstrap can be used to estimate the standard error using an empirical approach, rather than an explicit application of a formula.

The bootstrap method involves having a computer program generate a series of data sets that are designed to resemble the ones that would be observed if the estimation study were repeated many times. Each bootstrap data set is obtained by sampling (with replacement) from the original data. Call these data sets bootstrap samples.

For example, suppose N=8 and the following heights in inches are observed: Original sample: (68, 69, 69, 70, 71, 72, 72, 74). We use this example to make it very clear the steps involved in using bootstrap methods. The statistical validity of the bootstrap method is best when the sample size is 20 or more (Polansky, 1999).

Here are five bootstrap samples that were created by randomly sampling with replacement from the original sample (the values are ordered in magnitude to make the relative frequencies easier to see):

   Bootstrap sample 1: (68, 68, 68, 69, 69, 71, 72, 72)
   Bootstrap sample 2: (69, 69, 72, 72, 72, 72, 72, 74)
   Bootstrap sample 3: (69, 69, 69, 70, 71, 72, 72, 74)
   Bootstrap sample 4: (69, 70, 70, 72, 72, 72, 74, 74)
   Bootstrap sample 5: (69, 69, 69, 70, 71, 72, 72, 72)

Note that the original sample has two instances each of 69 and 72 inch subjects, and the values are more likely than any other to occur more than once in the bootstrap samples. However, this tendency is not a certainty. In sample 1, by chance the value 68 was chosen three times and the value 69 was chosen only once. Note also that the original sample does not contain someone 73 inches tall, and thus none of the bootstrap samples have a person of that height.

The mean of the original sample is 70.625 and the estimated standard error of the mean is .706. The means of the bootstrap samples are 69.63, 71.50, 70.75, 71.63, and 70.50. The standard deviation of these five values is .813, and this is the bootstrap estimated standard error, although we would never use a bootstrap estimate based on only five bootstrap samples.

The usual estimate of the confidence interval for the mean is 70.625±(.706)*(2.365) leading to the bounds (68.96, 72.29), where 2.365 is the *t* value corresponding to a tail area .975 and 7 degrees of freedom. As we expect, all five bootstrap sample means fall within this range. If we had constructed 100 bootstrap samples, however, we would have expected about five sample means to be either smaller than 68.98 or greater than 72.29. The confidence bound is expected to be accurate since it is well known that the sample mean is approximately normally distributed.

If we had created 1000 bootstrap samples, then we could have inferred the confidence region on the mean without calculating the standard error and assuming a normal distribution for the estimate. A <u>percentile estimate</u> of the 95% confidence interval is computed by ordering the 1000 bootstrap sample means from lowest to highest, and marking the 25th out of 1000 as the lower bound, and the 975th out of 1000 as the upper bound. Formally we call the ith ordered estimate $\theta^{(i)}$, and we choose the confidence bounds to be $\theta^{(\alpha)}$ and $\theta^{(1-\alpha)}$, where α is chosen so that total coverage is 1-2α. For the 95% CI, α is .025.

Although the percentile estimates are easy to calculate, it is known that they tend to be too narrow. Efron and Tibshirani (1993, pp 184-186) describe bias-corrected intervals that use bounds $\{\theta^{(\alpha 1)}, \theta^{(\alpha 2)}\}$ instead of the percentile bounds $\{\theta^{(\alpha)}, \theta^{(1-\alpha)}\}$. The values $\alpha 1$ and $\alpha 2$ are defined in a way that takes into account asymmetry in the distribution of bootstrap estimates. When the distribution of the bootstrap estimates is normal, the bias corrected interval and the percentile intervals are virtually the same.