Psychological Methods
2011, Vol. 16, No. 2, 93–115

1082-989X/11/$12.00 DOI: 10.1037/a0022658

Effect Size Measures for Mediation Models: Quantitative Strategies for Communicating Indirect Effects

Kristopher J. Preacher
University of Kansas

Ken Kelley
University of Notre Dame

The statistical analysis of mediation effects has become an indispensable tool for helping scientists investigate processes thought to be causal. Yet, in spite of many recent advances in the estimation and testing of mediation effects, little attention has been given to methods for communicating effect size and the practical importance of those effect sizes. Our goals in this article are to (a) outline some general desiderata for effect size measures, (b) describe current methods of expressing effect size and practical importance for mediation, (c) use the desiderata to evaluate these methods, and (d) develop new methods to communicate effect size in the context of mediation analysis. The first new effect size index we describe is a residual-based index that quantifies the amount of variance explained in both the mediator and the outcome. The second new effect size index quantifies the indirect effect as the proportion of the maximum possible indirect effect that could have been obtained, given the scales of the variables involved. We supplement our discussion by offering easy-to-use R tools for the numerical and visual communication of effect size for mediation effects.

Keywords: mediation, indirect effect, effect size

Supplemental materials: http://dx.doi.org/10.1037/a0022658.supp

Consider the case in which a researcher has established that some regressor (X) explains some of the variance in a criterion or dependent variable (Y) via regression. Equation 1 expresses the model for individual i:

$$Y_i = d_{Y.X} + cX_i + e_{Y.X_i}, \quad (1)$$

where c is the regression coefficient quantifying the *total effect* of X on Y, $d_{Y.X}$ is the intercept of the model, and $e_{Y.X_i}$ is the error associated with individual i. *Mediation analysis* consists of estimating the indirect effect of X on Y via an intervening variable called a mediator (M). In the simplest case, the researcher regresses M on X and separately regresses Y on both X and M using the following equations:

$$M_i = d_{M.X} + aX_i + e_{M.X_i}, \quad (2)$$

where $d_{M.X}$ is the intercept for M, a is the slope of M regressed on X, and $e_{M.X_i}$ is the error and

$$Y_i = d_{Y.MX} + bM_i + c'X_i + e_{Y.MX_i}, \quad (3)$$

where $d_{Y.MX}$ is the intercept for Y, b is the slope of Y regressed on M controlling for X, c' is the slope of Y regressed on X controlling for M, and $e_{Y.MX_i}$ is the error. The indirect effect, defined as $\hat{a} \times \hat{b}$, often is used as an index of mediation (where throughout a circumflex [ˆ] above a parameter denotes a sample estimate). In general, $\hat{a} \times \hat{b} = \hat{c} - \hat{c}'$, and thus $\hat{c} = \hat{a} \times \hat{b} + \hat{c}'$. Structural equation modeling may also be used to obtain both $\hat{a}$ and $\hat{b}$ simultaneously, correct for the attenuating effects of measurement error, and test more complex models, such as those where X, M, and Y are latent. Here we focus on the simplest case of a single mediator (unless otherwise stated) and no latent variables. Tests of mediation effects have become very popular in the managerial, behavioral, educational, and social sciences because they help researchers understand how, or by what means, effects unfold. A path diagram showing a simple mediation model is presented in Figure 1.

Many methods have been developed to facilitate significance testing and/or confidence interval formation for indirect effects (MacKinnon, 2008; MacKinnon, Lockwood, Hoffman, West, & Sheets, 2002). We find the increased attention being devoted to appropriate modeling and testing techniques highly encouraging. On the other hand, we believe this emphasis on modeling and statistical significance falls short of the ideal. Despite the recommendation of Baron and Kenny (1986, p. 1177) to consider the absolute size of relevant regression weights in addition to their statistical significance, very little attention has been devoted to quantifying and reporting the effect size of indirect effects in mediation models.

The fourfold purposes of this article are to (a) outline some general desiderata for effect size estimation, (b) review existing

This article was published Online First April 18, 2011.

Kristopher J. Preacher, Department of Psychology, University of Kansas; Ken Kelley, Department of Management, Mendoza College of Business, University of Notre Dame.

This study made use of the Socialization of Problem Behavior in Youth: 1969–1981 data made available by Richard and Shirley Jessor in 1991. The data are available through the Henry A. Murray Research Archive at Harvard University (producer and distributor). We thank Sonya K. Sterba, Scott E. Maxwell, and Robert Perera for valuable input.

Correspondence concerning this article should be addressed to Kristopher J. Preacher, Department of Psychology, University of Kansas, 1415 Jayhawk Boulevard, Room 426, Lawrence, KS 66045-7556. E-mail: preacher@ku.edu

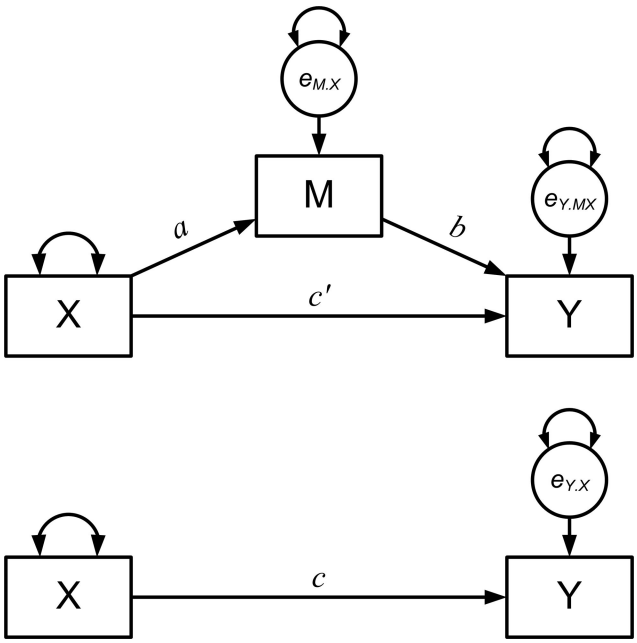

Figure 1. Diagram of models in which the effect of *X* on *Y* is (upper) versus is not (lower) mediated by *M*. Circles represent residuals, single-headed arrows represent regression weights, and double-headed arrows represent variance parameters.

effect sizes proposed in the mediation context, (c) use the desiderata to evaluate how effect size has been quantified and reported in the context of mediation, and (d) suggest new ways to communicate the magnitude of the indirect effect while avoiding the shortcomings of existing methods. The development of quality effect sizes will facilitate meta-analytic work on mediation, something currently lacking in the mediation literature. Finally, we provide R code (R Development Core Team, 2010) with the MBESS[1] package (Kelley & Lai, 2010; Kelley, 2007b) to aid researchers who wish to use the methods we describe in their own research. Graphical methods are an important supplement to quantitative descriptions of mediation and can themselves be useful ways of communicating results. We discuss graphical methods in an online supplement.[2]

Conceptualizing Effect Size: A Definition and Desiderata

Numerous methodologists have recommended that effect size measures accompany reports of statistical significance and nonsignificance. As a result, effect size reporting is now encouraged or mandated by many journal editors, as well as many organizations with scientific oversight, including the National Center for Education Statistics (NCES, 2003), the International Committee of Medical Journal Editors (via the Consolidated Standard of Reporting Trials [CONSORT; Moher et al., 2010]), and the American Educational Research Association (AERA, 2006). Furthermore, as the American Psychological Association (APA) Task Force on Statistical Inference recommended, reporting some measure of effect size is "essential to good research" and "enables readers to evaluate the stability of results across samples, designs, and analyses" (Wilkinson & the Task Force on Statistical Inference, 1999, p. 599). In addition, "it is almost always necessary to include some measure of effect size in the Results section" (American Psychological Association, 2010, p. 34). But even though researchers are now urged to report effect size to supplement or replace statistical significance, researchers who use mediation models have few resources to which to turn. For researchers who desire to report effect size for mediation effects, there simply is not much work that can be referenced (Albert, 2008; MacKinnon, Fairchild, & Fritz, 2007; Preacher & Hayes, 2008a), and many of the methods that do exist have limitations that often go unrecognized. We begin by offering a general definition of effect size, outlining some desirable properties (desiderata) to which new effect size measures should aspire, and delineating the issues that warrant attention when reporting effect size for mediation effects. Ultimately, we recommend a new effect size measure, developed in a later section, that we believe has desirable properties that will be useful in quantifying the magnitude of the indirect effect in the application of mediation models.

Defining Effect Size

There is almost universal agreement among methodologists that effect size is very important to report whenever possible (Grissom & Kim, 2005; Thompson, 2007; Vacha-Haase, Nilsson, Reetz, Lance, & Thompson, 2000). Yet, there are inconsistencies in how effect size is defined in the methodological literature, with the preponderance of authors favoring either a definition based on the magnitude of departure from a particular null hypothesis or a definition relating effect size to practical importance. For example, Cohen (1988) defined effect size as the "degree to which the phenomenon is present in the population or the degree to which the null hypothesis is false" (pp. 9–10). Similarly, Vacha-Haase and Thompson (2004) defined effect size as a "statistic that quantifies the degree to which sample results diverge from the expectations . . . specified in the null hypothesis" (p. 473). Other major works on effect size have similar definitions (Grissom & Kim, 2005). On the other hand, some authors prefer to regard effect size as any numeric quantity intended to convey the *practical significance* (or importance) of an effect (Kirk, 1996). Practical importance, in turn, is the substantive importance of an effect in real terms. That is, practical importance is the degree to which scientists, practitioners, executives, consumers, politicians, or the public at large, for example, would consider a finding important and worthy of attention. Yet other authors use both kinds of definition interchangeably (Henson, 2006). These two kinds of definitions—one based on the size of an effect relative to a null hypothesis and the other based on practical importance—imply related but separate concepts.

[1] Originally MBESS stood for Methods for the Behavioral, Educational, and Social Sciences. However, MBESS is now an orphaned acronym, meaning that what was an acronym is now literally its name.

[2] The supplemental material on graphical methods may be found at the *Psychological Methods* website and at the authors' websites (http://quantpsy.org and https://repository.library.nd.edu/view/5/Mediation_Effect_Sizes.pdf).

In response to the need for a general, inclusive definition of effect size, we define effect size as *any measure that reflects a quantity of interest, either in an absolute sense or as compared with some specified value*. The quantity of interest might refer to variability, association, difference, odds, rate, duration, discrepancy, proportionality, superiority, or degree of fit or misfit. It is possible for an effect size measure conforming to this definition to be used as an index of practical importance, although practical importance is not tied to our definition of effect size.

Desiderata for Good Effect Size Indices

Some desirable properties for effect size measures, which we term desiderata, are now outlined. First, virtually all effect size indices should be scaled appropriately, given the measurement and the question of interest. Without an interpretable scale, it is difficult to use effect size to communicate results in a meaningful and useful way. Often effect size is associated with standardized effect sizes; indeed, sometimes standardization is a defining characteristic of effect size, and in many cases, standardization frees the researcher from having to prepare a new set of interpretive benchmarks for every new scale or application (Cohen, 1988). Throughout, we define a standardized effect size as one that is not wedded to a particular measurement scale. More formally, it is an effect size that does not change in value based on linear transformations of the variable(s) involved. Although standardized effect sizes can be valuable, they are not always to be preferred over an effect size that is wedded to the original measurement scale, which may already be expressed in meaningful units that appropriately address the question of interest (Baguley, 2009; Frick, 1999). For example, group mean differences in scores on a widely understood instrument for measuring depressive symptoms are already expressed on a metric that is understandable to depression researchers, and to standardize effects involving the scale would only confuse matters.

Second, it should be emphasized that effect size estimates are themselves sample statistics and thus will almost certainly differ from their corresponding population values. Therefore, it is important to report confidence intervals for effect sizes because the real interest lies not in the estimated value but in the population value (Balluerka, Gómez, & Hidalgo, 2005; Bird, 2002; Cumming & Finch, 2001; Fidler & Thompson, 2001; Henson, 2006; Kelley, 2007a; Kirk, 1996; Smithson, 2001; Thompson, 2002, 2007). Third, although sampling error will affect the uncertainty in any effect size estimate and sampling error will tend to decrease as sample size (n) increases, the point estimate itself should be independent of sample size. Effect sizes are usually considered to have corresponding population values (parameters), so the estimation of an effect should be independent of the arbitrary size of the sample that is collected in order to estimate that population effect. Two researchers should not come to different conclusions about the size of an effect simply because their samples are of different sizes, all other things being equal. None of the effect sizes in common use depends on n for their respective definitions (r, Cohen's d, odds ratios, etc.) other than in a limited fashion that quickly diminishes as n increases. More broadly, the sample estimators of population effect sizes should be *unbiased* (i.e., the expected value of the effect size should equal the parameter over infinite repeated sampling), *consistent* (i.e., the effect size estimate should converge on the population value as n increases), and *efficient* (i.e., the effect size estimator should have reasonably low sampling variability).

Effect Size in the Context of Mediation Analysis

The magnitude of the indirect effect can be informally signified by the a and b coefficients themselves. MacKinnon (2008) and MacKinnon et al. (2007) suggested that either the standardized regression coefficient or the raw correlation can be used as an effect size measure for the a coefficient, and a partial correlation can be used as an effect size measure for the b coefficient. This method is not entirely satisfactory, as a and b alone do not convey the full meaning of an indirect effect. Therefore, it is important to develop a way to gauge the effect size of the product term ab itself. Unfortunately, the indirect effect does not fit any of the classic effect size measures developed in methodological works or reported in research, such as the standardized mean difference (Cohen's d, Hedges' g), association (β, r, r_{bis}), odds ratio (OR), percentage of variance explained (intraclass correlation, R^2, η^2, ω^2), or the coefficient of variation. In mediation models, the primary effect of interest is an indirect effect. Such an effect is complex because it is the product of (here) two regression coefficients and does not fit conveniently into the framework of existing effect sizes. Thus, it is challenging to adapt existing effect size measures for use in mediation analysis. In developing and evaluating new methods of expressing effect size for indirect effects, it will be important to do so in light of the definition and desiderata outlined earlier. That is, effect sizes suggested for mediation analysis should be on a meaningful metric, should be amenable to the construction of confidence intervals, and should be independent of sample size. A meaningful metric in this context is any metric where the size of the effect can be interpreted in a meaningful way vis-à-vis the constructs under study. Standardized effect sizes are on a meaningful scale in units of standard deviations. For example, in a regression model with a single independent variable and a single dependent variable that are both standardized, a correlation coefficient can be interpreted as the number of standard deviations that the dependent variable is expected to increase for a change of one standard deviation in the independent variable. Our suggestion for effect sizes to be on a meaningful metric implies no preference for standardized or unstandardized effect sizes. The metric that most effectively communicates the particular effect size in the specific context is what we regard as the preferred metric. This will vary by situation.

Illustrative Example

To make our discussion more concrete, we make use of a publicly available data set, Jessor and Jessor's (1991) *Socialization of Problem Behavior in Youth 1969–1981* (SPBY; Jessor & Jessor, 1991). The sample size is $n = 432$ with complete data. In the applications to follow, the predictor variable is *achievement values* (VAC), obtained by averaging 10 items from the Personal Values Questionnaire (Jessor & Jessor, 1977) administered in 1969 to high school students in the Boulder area of Colorado. Example items ask respondents how much they like having good grades for entering college and how much they like being on the honor roll.

The mediator variable is *attitude toward deviance* (ATD), obtained by averaging 30 items from the Attitude Toward Deviance Scale (Jessor & Jessor, 1977) administered to the same students in 1970. Example items ask respondents how wrong it is to break into a locked place or to beat up another kid. Because of the manner in which responses were scored, higher scores on ATD indicate greater intolerance of deviant behavior. The outcome variable is *deviant behavior* (DVB), obtained by averaging 30 items from the Deviant Behavior Report Scale (Jessor & Jessor, 1977) administered to the same sample in 1971. An example item asks respondents how often they have threatened a teacher out of anger. Basic results for the direct and indirect effects linking VAC, ATD, and DVB are provided in Table 1, and covariances, correlations, and means for the three variables are provided in Table 2. Figure 2 is a Venn diagram depicting the variances of VAC, ATD, and DVB as circles, overlapping to the degree that these variables are related.

Table 2
Correlations, Covariances, and Means for Jessor and Jessor's (1991) Data

	VAC (X)	ATD (M)	DVB (Y)
VAC (X)	2.268	*.291*	*−.190*
ATD (M)	0.662	2.276	*−.493*
DVB (Y)	−0.087	−0.226	0.092
M	7.158	5.893	1.649

Note. Numbers on the diagonal are variances, those below the diagonal are covariances, and those above the diagonal (italicized) are correlations. VAC = (higher) achievement values; ATD = (more intolerant) attitude toward deviance; DVB = (more) deviant behavior.

Existing Methods of Expressing Effect Size for Mediation Effects

In this section, we describe and evaluate existing measures of effect size for mediation effects. Each method is evaluated in light of the definition and desiderata identified above and illustrated using SPBY data.

Verbal Descriptors

The literature about, and using, mediation is fraught with language invoking the idea of effect size but not directly addressing it in a rigorous, quantitative manner. The most popular way to express effect size for mediation is through informal descriptors, such as *complete*, *perfect*, or *partial* mediation (Mathieu & Taylor, 2006). James and Brett (1984) described *complete mediation* as occurring when the effect of X on Y completely disappears (i.e., $c' = 0$) when M is added as a predictor of Y. Baron and Kenny (1986) asserted that "the strongest demonstration of mediation occur[s] when Path [c'] is zero" (p. 1176), effectively proposing a way to judge the effect size of an indirect effect by examining the statistical significance of c'. The condition in which $c' = 0$ after the detection of a statistically significant mediation effect they dub *perfect mediation* (p. 1177). In practice, a researcher may claim that a mediation effect is *perfect* or *complete* if c' is not statistically significantly different from zero, which is to say that perfect mediation exists when there is not sufficient evidence to demonstrate that it does not. In other words, the status quo is to claim perfect mediation when the null hypothesis that $c' = 0$ is not rejected by the null hypothesis significance test, thus using the absence of evidence (i.e., a failure to reject the null hypothesis that $c' = 0$) as evidence of absence (of the direct effect exerted by X on Y). For example, in the SPBY data, $c' = -.0102$ ($p = .25$, *ns*; 95% CI [−.028, .007]), and thus the statistically significant indirect effect would signify complete mediation by Baron and Kenny's criterion. Of course, one could fail to reject the null hypothesis that $c' = 0$ due to insufficient statistical power from an insufficiently large n. Furthermore, it is not clear what should be done when $c' < 0$ by

Table 1
Regression Results for the Mediation of the Effect of Achievement Values on Deviant Behavior by Attitude Toward Deviance

Model	Estimate	*SE*	*p*	CI (lower)	CI (upper)
Model without mediator					
Intercept	1.9236	.0698	<.0001	1.7864	2.0608
VAC → DVB (c)	−.0383	.0095	.0001	−0.0571	−0.0196
$R^2_{Y,X}$	.0361			0.0095	0.0779
Model with mediator					
Intercept	2.2900	.0704	<.0001	2.1517	2.4282
VAC → ATD (a)	.2916	.0462	<.0001	0.2008	0.3825
ATD → DVB (b)	−.0963	.0088	<.0001	−0.1136	−0.0789
VAC → DVB (c')	−.0102	.0088	.2472	−0.0276	0.0071
Indirect effect ($a \times b$)	−.0281			−0.0390	−0.0189
$R^2_{M,X}$	.0848			0.0408	0.1405
$R^2_{Y,MX}$	.2456			0.1750	0.3155

Note. Regression weights a, b, c, and c' are illustrated in Figure 1. $R^2_{Y,X}$ is the proportion of variance in Y explained by X, $R^2_{M,X}$ is the proportion of variance in M explained by X, and $R^2_{Y,MX}$ is the proportion of variance in Y explained by X and M. The 95% CI for $a \times b$ is obtained by the bias-corrected bootstrap with 10,000 resamples. The CIs for R^2 indices are obtained analytically. In this example, VAC (achievement values) is the independent variable (X), ATD (attitude toward deviance) is the mediator (M), and DVB (deviant behavior) is the outcome (Y). CI (lower) = lower bound of a 95% confidence interval; CI (upper) = upper bound; → = affects.

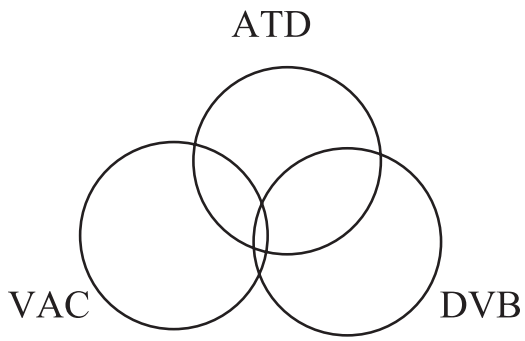

Figure 2. Venn diagram showing the extent to which VAC, ATD, and DVB share variance in common. Each circle represents the total variance of a variable, and the overlap of two circles represents the portion of variance shared in common by two variables. VAC = (higher) achievement values; ATD = (more intolerant) attitude toward deviance; DVB = (more) deviant behavior.

even a small amount. Baron and Kenny cautioned that at least in psychology, complete mediation is expected to be rare because of the prevalence of multiple mediators. These descriptors are found in common usage and are intended to denote either the practical importance of an effect (describing an effect as *complete* carries the implication that it is "large" or "important," whereas, a *partial* mediation effect is not as impressive) or the potential for identifying additional mediators (*complete* implies that there is no room for further mediators, whereas *partial* potentially indicates a need to continue looking for additional mediators).

The informal descriptors *complete* and *partial* do not fulfill the desiderata identified earlier. First, they are not expressed in a meaningfully scaled metric. Although the words *complete* and *partial* invoke the idea of proportion, they are not numerical, so the importance attached to the terms is largely subjective. Second, because they are not numerical, it is impossible to compute confidence intervals for them. Third, these descriptors are defined in terms of the statistical significance of c' and so are not independent of sample size. Because of this, we argue that a researcher is implicitly rewarded for using a small sample with a greater likelihood of obtaining "complete mediation," which runs counter to the universal recommendation to prefer larger samples. Fourth, although they do convey something about practical importance, they are highly imprecise. In general, holding everything else constant, it is more likely that a mediator will completely mediate a relatively small total effect (c) than a relatively large total effect, so an effect in which M partially mediates a relatively large c may be more impressive than one in which M completely mediates a relatively small c.

Ratio Measures of Relative Magnitude

Several quantitative measures of *relative magnitude*, in addition to the verbal descriptors discussed earlier, have been proposed for mediation effects. Alwin and Hauser (1975) proposed several such measures in their classic article on the decomposition of effects in path analysis (see also MacKinnon, 1994; MacKinnon & Dwyer, 1993; Sobel, 1982). Two measures that are relevant for simple mediation models are the ratio of the indirect effect to the total effect,

$$P_M = \frac{ab}{ab + c'} = \frac{ab}{c} = 1 - \frac{c'}{c}, \tag{4}$$

and the ratio of the direct effect to the total effect,

$$1 - P_M = 1 - \frac{ab}{ab + c'} = 1 - \frac{ab}{c} = \frac{c'}{c}, \tag{5}$$

where a is the slope linking X to M, b is the conditional slope linking M to Y, c is the *total effect* of X on Y, and c' is the conditional slope linking X to Y (Alwin & Hauser, 1975; Buyse & Molenberghs, 1998; MacKinnon, 2008; MacKinnon et al., 2007; MacKinnon, Warsi, & Dwyer, 1995; Shrout & Bolger, 2002; Tofighi, MacKinnon, & Yoon, 2009; Wang & Taylor, 2002). The sample statistic $\hat{P}_M$ is obtained by substituting sample quantities for their corresponding population values. P_M is also known as the *validation ratio* (Freedman, 2001) or *mediation ratio* (Ditlevsen, Christensen, Lynch, Damsgaard, & Keiding, 2005) in epidemiological research and as the *relative indirect effect* (Huang, Sivaganesan, Succop, & Goodman, 2004) and is often interpreted loosely as the proportion of the total effect that is mediated. In the SPBY data, $\hat{P}_M = \frac{\hat{a}\hat{b}}{\hat{a}\hat{b} + \hat{c}'} = \frac{(.2916)(-.0963)}{(.2916)(-.0963) - .0102} = .733$ (95% CI [.458, 1.357]),[3] signifying, if $\hat{P}_M$ is to be interpreted as a proportion (an assumption we soon question), that attitudes toward deviance mediate approximately three-fourths of the total effect of achievement values on deviant behavior. The complement of $\hat{P}_M$, if in fact it is interpreted as a proportion, is thus $1 - \hat{P}_M = .266$.

Sobel (1982) proposed the ratio of the indirect effect to the direct effect:

$$R_M = \frac{ab}{c'}. \tag{6}$$

A recent example of the use of R_M is provided by Barreto and Ellemers (2005), who reported that the ratio of the indirect to direct effect of type of sexism (hostile vs. benevolent) on perceived sexism through evaluation of the source was 1.7. In the SPBY data, $\hat{R}_M = \frac{\hat{a}\hat{b}}{\hat{c}'} = \frac{(.2916)(-.0963)}{-.0102} = 2.742$ (95% CI [−4.162, 147.689]), indicating that the indirect effect of VAC on DVB is approximately 2.75 times the size of the direct effect, but this ratio is not statistically significantly different from zero at the 5% level because 0 is contained in the 95% confidence interval.

Although P_M and R_M are easy to estimate in samples, as measures of effect size they suffer from several limitations; we discuss limitations of P_M first, followed by limitations of R_M. First, consider that as an index P_M can convey misleading estimates of practical importance. Depending on the context, obtaining $P_M = .9$ for a relatively small but statistically significant total effect may not necessarily be as impressive as obtaining $P_M = .6$ for a relatively large and statistically significant total effect, yet the former sounds as if it is somehow more important, whereas the latter seems as though it is less impressive when quantified using a standardized effect size like $\hat{P}_M$. As we discuss later, it is important to be mindful of the distinction between the value of an effect size, even if it seems rather small or large, and the practical importance of the effect size in the specific context. Second, despite the fact that many researchers refer to it as a proportion, $\hat{P}_M$ is not a proportion and thus cannot be interpreted as such. The quantity $\hat{a}\hat{b}/(\hat{a}\hat{b} + \hat{c}')$ can exceed 1.0 or be negative, depending on the relation of $\hat{c}'$ to $\hat{c}$ (Albert, 2008; Alwin & Hauser, 1975; MacKinnon, 2008), which implies that it is not a proportion. The fact that

[3] This and subsequently reported confidence intervals use bias-corrected and accelerated (BCa) bootstrap confidence limits. Bootstrapping involves treating the original sample as if it were a population and simulating the sampling process assumed to have led to the original sample. An arbitrarily large number B of bootstrap samples of size n are selected with replacement from the original sample of size n. (B is recommended to be several thousand for acceptable precision; we used $B = 10{,}000$.) Each of these B "resamples" is used to compute the statistic of interest, resulting in B bootstrap estimates of the statistic. The empirical sampling distribution of these bootstrap estimates serves as a basis for obtaining confidence limits by referring to values at the appropriate percentiles (e.g., 2.5 & 97.5) for what are termed percentile confidence intervals. BCa confidence limits are obtained by adjusting the limits from the percentile confidence intervals according to instructions provided by Efron (1987) and Efron and Tibshirani (1993).

P_M is not literally a proportion is not a limitation of P_M per se but rather of how P_M has been discussed and used. Nevertheless, since P_M cannot be appropriately interpreted as a proportion, it is less useful than its label implies. Measures of explained variance are better suited to bear such proportion interpretations, which we discuss later. Third, focusing on the overall value of P_M may neglect additional mediators in models where multiple mediators are plausible (MacKinnon et al., 2007). It is easy to assume that if $\hat{P}_M$ seems large (i.e., approaches 1.0, which, as we indicated earlier, is not its upper limit), there is "no room" for additional mediators, when in fact it is possible to identify additional and/or better mediators. An additional mediator may well be correlated with the one already included in the model, in which case the indirect effect would be partitioned into parts unique to each mediator. Fourth, $\hat{P}_M$ and $\hat{R}_M$ have large variances over repeated samples, and thus they are not very efficient estimators. In fact, MacKinnon (1994) showed that both ratio measures can be unstable and commented that they "should be used only with relatively large sample sizes" (p. 139). Simulation research has shown that $\hat{P}_M$ is unstable unless $n > 500$ (MacKinnon, 1994; MacKinnon et al., 1995). Similarly, $\hat{R}_M$ is unstable unless $n > 5{,}000$ (MacKinnon et al., 1995). R_M is so unstable because the numerator (ab) varies inversely with the denominator (c'). Consequently, minor fluctuations in ab and c' can lead to large fluctuations in their ratio. These large fluctuations can become enormous when c' is near zero because R_M approaches infinity when c' approaches zero. Examination of Figure 3 shows this sensitivity for specific situations, where the value of R_M abruptly approaches positive or negative infinity as the value of c is approached. Tofighi et al. (2009) similarly reported that very large samples are required for stable estimation of ratio measures. Both of these measures vary in bias and precision as a function of the size of the effects, with larger effects imparting less bias and being more precise. Taken together, these four limitations make us question the usefulness of P_M as a population value worth estimating and interpreting.

Although the ratio measure R_M does not have any pretensions toward being a proportion, it simply repackages the same information as P_M without conveying any additional information [$R_M = P_M/(1 - P_M)$]. Like P_M, R_M can assume values that exaggerate relatively small effects or trivialize relatively large ones. Considering the reasonable case where $\hat{c} = .63$ and $\hat{c}' = -.01$, the ratio of the indirect to direct effect will equal a nonsensical -64, yet if $\hat{c} = .63$ and $\hat{c}' = +.01$, the ratio will equal $+62$. In addition, if $\hat{c}$ is relatively small but $\hat{a}\hat{b}$ is relatively large, the ratio can assume extremely large values, as R_M is an unbounded quantity. Conversely, if $\hat{c}$ is relatively large and $\hat{a}\hat{b}$ is relatively small, small yet substantively important effects can easily slip through the cracks. Figure 3 shows that for a fixed value of the total effect ($c = .4$), R_M assumes small values for most indirect effects likely to occur between .0 and .35 and then increases rapidly to $+\infty$ as c is approached from below. For indirect effects above c, R_M approaches $-\infty$ as c is approached from above.

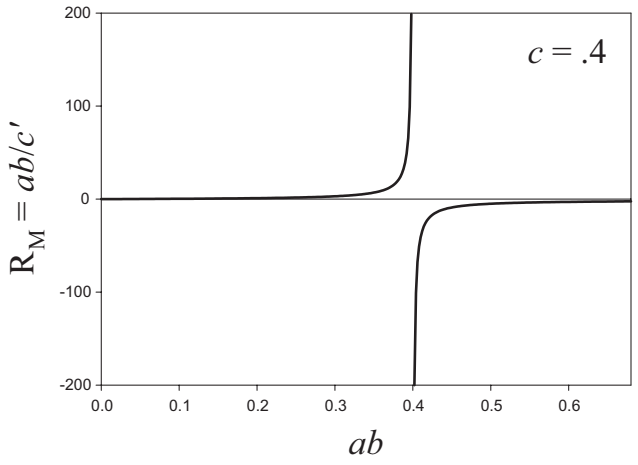

Figure 3. Plot of the ratio measure R_M for total effect $c = .4$ and indirect effects ab ranging from 0 to .7.

Although the limitations of $\hat{P}_M$ and $\hat{R}_M$ we note above are serious, estimates $\hat{P}_M$ and $\hat{R}_M$ are currently the most widely used measures of effect size. There are perhaps four reasons why $\hat{P}_M$ and $\hat{R}_M$ are so widely used. First, consistent with our third desideratum, the estimates $\hat{P}_M$ and $\hat{R}_M$ are relatively unaffected by sample size. Second, Alwin and Hauser (1975) noted that the proportionate decomposition of effects into direct and indirect components can facilitate interpopulation comparison of such effects, even when the variables of interest are not measured on the same scales across groups. Third, consistent with our second desideratum, both P_M and R_M are amenable to the construction of confidence intervals. Regarding confidence interval construction, Lin, Fleming, and DeGruttola (1997) gave a confidence interval for P_M based on the delta method, and this interval and one based on Fieller's method are discussed by Freedman (2001) and Wang and Taylor (2002).[4] Sobel (1982) provided derivations necessary for constructing an asymptotic confidence interval for $\hat{R}_M$. MacKinnon et al. (1995) and Tofighi et al. (2009) provided delta method standard errors for both ratio measures, for the cases where b and c' are either correlated or uncorrelated. However, because neither $\hat{P}_M$ nor $\hat{R}_M$ is normally distributed except in very large samples, it is not advisable to use any of the above noted confidence interval methods but rather to use the bootstrap approach, as we have discussed (e.g., Wang & Taylor, 2002). The fourth reason that $\hat{P}_M$ and $\hat{R}_M$ are so widely used, we believe, is that there really have been no better alternatives proposed in the literature for communicating the magnitude of effect.

Buyse and Molenberghs (1998) suggested a ratio that we abbreviate as S_M:

$$S_M = \frac{c}{a}. \tag{7}$$

[4] The delta method is used to derive an approximate probability distribution for a function $g(\hat{\boldsymbol{\theta}})$ of asymptotically normal parameter estimates in the vector $\hat{\boldsymbol{\theta}}$. It proceeds by first finding the first- or second-order (usually higher orders are not necessary) Taylor series expansion of the function, $\tilde{g}(\hat{\boldsymbol{\theta}})$, and then applying the definition of a variance, $\text{var}(\tilde{g}(\hat{\boldsymbol{\theta}})) = E[(\tilde{g}(\hat{\boldsymbol{\theta}}))^2] - [E(\tilde{g}(\hat{\boldsymbol{\theta}}))]^2$. The delta method is commonly used to derive estimated standard errors for functions of parameter estimates, which can then be used to construct confidence intervals for estimates that are assumed normally distributed. Fieller's method involves linearizing the ratio, finding the values of the squared linearized form that are less than or equal to the desired critical value under the χ^2 distribution, then solving for values of $g(\hat{\boldsymbol{\theta}})$ satisfying the inequality.

S_M is a measure of the success of a surrogate endpoint, a measure of an intermediate variable that may be related to an important clinical endpoint. For example, gum inflammation may be treated as a surrogate endpoint for tooth loss, and LDL cholesterol is often treated as a surrogate for heart disease. Thus, although surrogate endpoints share much in common with mediators, the emphasis is on coefficients *a* and *c* rather than *a* and *b*. The ratio *c*/*a* should be about 1.0 if *X* predicts *M* to the same extent that it predicts *Y* (MacKinnon, 2008; Tofighi et al., 2009). Tofighi et al. (2009) provided a delta method standard error for this ratio measure and recommend a sample size of at least 500 for accurate *SE*s when the regression weights are small. In the SPBY data, $\hat{S}_M = \frac{\hat{c}}{\hat{a}} = \frac{-.0383}{.2916} = -.131$ (95% CI [−.195, −.077]). However, we caution that S_M has at least two flaws that limit its usefulness as an effect size measure for mediation. First, it does not incorporate *b*, a crucial component of the indirect effect. Thus, the indirect effect could be quite small or even zero for even a respectably sized $\hat{S}_M$. Second, because it is a ratio, S_M depends on the relative size of the component parameters rather than their absolute magnitudes. As an example of why this might be problematic, consider the case of standardized coefficients $\hat{a} = .0001$ and $\hat{c} = .0001$. In a situation in which $\hat{c} = .0001$ is a trivial effect (even if statistically significant), we probably should not be impressed by $\hat{S}_M = 1$.

Unstandardized Indirect Effect

It often is not appreciated that statistics in their original metrics can be considered effect sizes if they are directly interpretable (Abelson, 1995; Baguley, 2009; Frick, 1999; Ozer, 2007). The most obvious method of expressing the magnitude of the indirect effect is to directly interpret the sample $\hat{a}\hat{b}$ as an estimate of the population *ab*. The unstandardized indirect effect $\hat{a}\hat{b}$ is independent of *n* and can be interpreted using the original scales of the variables in the model. The product *ab* has a straightforward interpretation as the decrease in the effect of *X* on *Y* when *M* is added to the model or as the amount by which *Y* is expected to increase indirectly through *M* per a unit change in *X*. In the SPBY data, for example, $\hat{a}\hat{b} = -.0281$ (95% CI [−.039, −.019]), implying that DVB is expected to decrease by .0281 units (on its 4-point scale) for every one-unit increase in VAC (on its 10-point scale) if one considers only the indirect influence via ATD.

If the variables *X* and *Y* are already on meaningful metrics, simply reporting *ab* and interpreting it may suffice to communicate effect size and practical importance. As has been discussed in the mediation literature, there are multiple ways to construct confidence intervals for *ab*, the product term does not depend on *n*, and the product conveys information about practical importance if the units of *X* and *Y* bear meaningful interpretation. If, however, the metric of either *X* or *Y* (or both) is arbitrary (as is the case in much applied work), not easily interpretable, or not well calibrated to the phenomenon of interest, it may not be sensible to directly interpret *ab*. Without knowing more about the scales of VAC and DVB, how they are applied in certain areas, or what should be considered "impressive" in the specific context of predicting deviant behavior using the Deviant Behavior Report Scale, it is difficult to know whether to be impressed by the finding that DVB is expected to decrease by .0281 units per unit change in VAC indirectly through ATD. A disadvantage of using *ab* as an effect size measure is that it is not robust to changes in scale, which limits its usefulness in meta-analysis.

Partially Standardized Indirect Effect

MacKinnon (2008) suggested that indirect effects may be standardized in the following way:

$$ab_{ps} = \frac{ab}{\sigma_Y}, \tag{8}$$

which is the ratio of the indirect effect to the standard deviation of *Y*. This index represents the size of the indirect effect in terms of standard deviation units in *Y*. Because *ab* is interpreted in raw units of *Y*, dividing by σ_Y removes the scale of *Y*, leaving a metric standardized in *Y* but not *X* or *M*. The interpretation of ab_{ps} is the number of standard deviations by which *Y* is expected to increase or decrease per a change in *M* of size *a*. Coefficient *a*, in turn, remains unstandardized. In the SPBY example, $\hat{a}\hat{b}_{ps} = \frac{\hat{a}\hat{b}}{s_Y} = \frac{(.2916)(-.0963)}{.3036} = -.092$ (95% CI [−.125, −.064]), implying that DVB is expected to decrease by .092 standard deviations for every one-unit increase in VAC (on its 10-point scale) indirectly via ATD.

Completely Standardized Indirect Effect

Carrying MacKinnon's (2008) logic further, we could fully standardize the indirect effect by multiplying ab_{ps} by s_X. The resulting index would be fully insensitive to the scales of *X*, *M*, and *Y*. Preacher and Hayes (2008a) suggested the term *index of mediation* for this effect size measure:

$$ab_{cs} = ab\frac{\sigma_X}{\sigma_Y}. \tag{9}$$

Alwin and Hauser (1975, p. 41) and Cheung (2009) discussed this index as well, noting that it can be used to compare indirect effects across populations or studies when variables use different metrics in each population. Thus, standardized indirect effects may be useful in meta-analysis. However, as we note later, many authors point out that the standardization factor varies from study to study, implying that standardized effect sizes may be less useful than is generally thought. Bobko and Rieck (1980) also considered indirect effects using standardized variables, and Raykov, Brennan, Reinhardt, and Horowitz (2008) advocated a scale-free correlation structure modeling approach to estimating mediation effects. In the SPBY example, $\hat{a}\hat{b}_{cs} = \hat{a}\hat{b}\frac{s_X}{s_Y} = (.2916)(-.0963)\frac{1.5061}{.3036} = -.139$ (95% CI [−.187, −.097]), indicating that DVB decreases by .139 standard deviations for every 1 *SD* increase in VAC indirectly via ATD.

To summarize the three effect size measures just described, note that all three may be expressed in terms of standardized regression weights (β) and standard deviations:

$$ab = \beta_{MX}\beta_{YM}\left(\frac{\sigma_Y}{\sigma_X}\right); \tag{10}$$

$$ab_{ps} = \beta_{MX}\beta_{YM}\left(\frac{1}{\sigma_X}\right); \quad (11)$$

$$ab_{cs} = \beta_{MX}\beta_{YM}. \quad (12)$$

It is interesting to note that the metric of M is absent from all three indices. The formula for coefficient a includes a $\left(\frac{\sigma_M}{\sigma_X}\right)$ term, and the formula for b includes a $\left(\frac{\sigma_Y}{\sigma_M}\right)$ term; the σ_M terms cancel when a and b are multiplied (MacKinnon, 2000; Preacher & Hayes, 2008b). It is a simple matter to construct confidence intervals for any of these indices (the bootstrap is recommended; Cheung, 2009), and none of them depend on sample size. Even though ab_{ps} is partially standardized, the fact that it relies in part on the metric of X prevents it from being used to compare indirect effects across multiple studies, even though it can be used to quantify effect size for a given study if the scale of X can be meaningfully interpreted. Of the three indices above, only ab_{cs} can generally be used in other situations where it is important to compare indirect effects across situations using different metrics for X and/or Y. A possible limitation of ab_{cs} is that it is not bounded in the way that a correlation or a proportion is—either component may be negative, and β_{YM} may exceed 1.0. Nevertheless, unlike P_M, ab_{cs} retains its interpretability when this happens.

On the other hand, not all methodologists support the use of standardized effect sizes. Bond, Wiitala, and Richard (2003), for example, strongly cautioned against the use of standardized mean differences in meta-analysis. Achen (1977), Baguley (2009), Greenland (1998), Greenland, Schlesselman, and Criqui (1986), Kim and Ferree (1981), King (1986), and O'Grady (1982) are decidedly pessimistic about the use of correlations and r^2 and other standardized effect sizes for expressing effects, as they depend on the variances of the measured variables.

Indices of Explained Variance

A common type of effect size is expressed in terms of explained variance. That is, the researcher often seeks to include predictors of a criterion such that the variance of residuals is reduced by some nontrivial amount. For example, η^2 and ω^2 in the analysis of variance framework, intraclass correlation in the mixed-model framework, and R^2 in the regression framework all can be interpreted as proportions of explained variance. These indices equate effect size with the proportion of the total variance in one variable shared with, or explained by, one or more other variables. They are popular as effect size estimates in part because they use an easily interpretable standardized metric, namely, a proportion metric. Therefore, it is not surprising that such measures should be considered in the mediation context as well.

MacKinnon (2008) suggested three such measures for use in the mediation context. Here they are referred to by his equation numbers (4.5, 4.6, and 4.7) to distinguish among them.

$$R^2_{4.5} = r^2_{YM} - (R^2_{Y,MX} - r^2_{YX}); \quad (13)$$

$$R^2_{4.6} = (r^2_{MX})(r^2_{YM.X}); \quad (14)$$

$$R^2_{4.7} = \frac{(r^2_{MX})(r^2_{YM.X})}{R^2_{Y,MX}}. \quad (15)$$

The $R^2_{Y,MX}$ term in the expressions for $R^2_{4.5}$ and $R^2_{4.7}$ is the proportion of variance in Y together explained by X and M; visually, $R^2_{Y,MX}$ corresponds to the proportion of the DVB circle in Figure 2 that is also covered by the VAC or ATD circles. The term r^2_{YX} is the squared correlation of X and Y (the proportion of the DVB circle occluded by VAC), and $r^2_{YM.X}$ is the squared partial correlation of Y with M, partialling out X (the proportion of the DVB circle not shared with VAC that *is* shared with ATD). Alternative expressions yielding each of these indices purely in terms of multiple R^2 (for ease of computation) are

$$R^2_{4.5} = R^2_{Y,M} - (R^2_{Y,MX} - R^2_{Y,X}); \quad (13b)$$

$$R^2_{4.6} = \frac{R^2_{M,X}(R^2_{Y,MX} - R^2_{Y,X})}{1 - R^2_{Y,X}}; \quad (14b)$$

$$R^2_{4.7} = \frac{R^2_{M,X}(R^2_{Y,MX} - R^2_{Y,X})}{R^2_{Y,MX}(1 - R^2_{Y,X})}. \quad (15b)$$

An equivalent expression for $R^2_{4.5}$ is

$$R^2_{4.5} = r^2_{YM} - r^2_{Y(M.X)}, \quad (16)$$

where $r^2_{Y(M.X)}$ is the squared semipartial correlation of Y with the part of M from which X has been partialed. $R^2_{4.5}$ has a straightforward interpretation as the overlap of the variances of X and Y that also overlaps with the variance of M, or "the variance in Y that is common to both X and M but that can be attributed to neither alone" (Fairchild, MacKinnon, Taborga, & Taylor, 2009, p. 488). Overall, $R^2_{4.5}$ has many of the characteristics of a good effect size measure: (a) It increases as the indirect effect approaches the total effect c and so conveys information useful in judging practical importance; (b) it does not depend on sample size; and (c) it is possible to form a confidence interval for the population value. In the SPBY data example, $R^2_{4.5} = r^2_{YM} - (R^2_{Y,MX} - r^2_{YX}) = (-.4932)^2 - (.2456 - (-.1901)^2) = .034$ (95% CI [.010, .064]). In some situations, $R^2_{4.5}$ can be negative, as it is not literally the square of another value. Fairchild et al. (2009) noted that a negative $R^2_{4.5}$ can indicate that suppression rather than mediation is occurring. However, because negative values can occur, $R^2_{4.5}$ is not technically a proportion of variance as the label R^2 would seem to imply (Fairchild et al., 2009). We believe this limits the usefulness of $R^2_{4.5}$ as an effect size, but we do not rule out that it may have heuristic value in certain situations.

Unlike $R^2_{4.5}$, $R^2_{4.6}$ is a product of two squared correlations, in this case the squared correlation between X and M and the squared partial correlation of M and Y, partialling for X. In other words, $R^2_{4.6}$ is the proportion of Y variance that is *not* associated with X but *is* associated with M, weighted by the proportion of variance explained in M by X. Like $R^2_{4.5}$, it increases roughly as the indirect effect increases. Like $R^2_{4.5}$, it is standardized and does not depend on n, and it is possible to form confidence intervals for it. However, even though the lower bound is 0, and

it cannot exceed 1,[5] $R^2_{4.6}$ is difficult to interpret because it is the product of two proportions of variance. Because it is the product of two R^2 measures that are computed for different variables, it is not itself a proportion of variance as the label R^2 would imply.[6] Therefore, it is not appropriate to interpret it on an R^2 metric. Of the three R^2 indices suggested by MacKinnon, $R^2_{4.6}$ bears the closest resemblance to *ab*, and regardless of its interpretability as a proportion, it mirrors effect size very well. In the SPBY data example, $R^2_{4.6} = (r^2_{MX})(r^2_{YM.X}) = (.2911)^2(-.4662)^2 = .018$ (95% CI [.009, .032]). That is, .018 is the proportion of variance in deviant behavior that is not associated with achievement values but is associated with attitude toward deviance, weighted by the proportion of variance in attitude toward deviance explained by achievement values.

$R^2_{4.7}$ is simply $R^2_{4.6}$ divided by $R^2_{Y,MX}$, the proportion of variance in *Y* together explained by *X* and *M*. Because it divides by a number that is between 0 and 1, $R^2_{4.7}$ represents a simple rescaling of $R^2_{4.6}$. Correspondingly, we find $R^2_{4.7}$ difficult to interpret. Whereas it is bounded from below by 0, it can exceed 1, but not in situations likely to correspond to mediation. Because of this, it (like the other two R^2 indices) cannot be interpreted on a standardized proportion metric. In the SPBY example, $R^2_{4.7} = (r^2_{MX})(r^2_{YM.X})/R^2_{Y,MX} = .0184/.2456 = .075$ (95% CI [.041, .119]).

We present plots to enable readers to anticipate the behavior of various R^2 statistics. We do not suggest that similar figures be produced in applied research. These figures are intended to help readers better understand the ranges that the values can assume. Each plot was created by generating 15,000 random 3 × 3 correlation matrices,[7] denoted **R**; fitting a simple mediation model to each **R**; and plotting relevant statistics and effect size indices. For example, Figure 4 displays plots of $R^2_{4.5}$ plotted against *ab* for 15,000 randomly generated negative and indirect effects, holding the standardized total effect *c* constant at .2 (top) and .8 (bottom). From Figure 4 we can tell that when *c* is held constant, the most extreme positive score of $R^2_{4.5}$ is c^2. The effect size cannot exceed the square of the standardized total effect. $R^2_{4.6}$ is plotted as a function of *ab* in Figure 5 for 15,000 randomly generated indirect effects, holding the standardized total effect *c* constant at .2 (top) and .8 (bottom). $R^2_{4.7}$ is plotted as a function of *ab* in Figure 6 for 15,000 randomly generated indirect effects, holding the standardized total effect *c* constant at .2 (top) and .8 (bottom).

A related index was suggested by Lindenberger and Pötter (1998). Their *shared over simple effects* (*SOS*) index is the ratio of the variance in *Y* explained by both *X* and *M* divided by the variance in *Y* explained by *X*:

$$SOS = \frac{1}{r^2_{YX}}[r^2_{YM} - (1 - r^2_{YX})r^2_{YM.X}], \quad (17)$$

where $r^2_{YM.X}$ is the partial correlation of *M* and *Y* after partialling out *X*. A simpler expression for *SOS* in terms of indices already presented is

$$SOS = \frac{R^2_{4.5}}{r^2_{YX}}. \quad (18)$$

The authors describe *SOS* as the proportion of *X*-related variance in *Y* that is shared with *M*. Positive values of *SOS* indicate mediation, a value of 0 indicates no indirect effect, and negative values indicate suppression. In the SPBY example, *SOS* = .034/.036 = .934 (95% CI [.727, .999]). Because *SOS* can assume values less than zero or greater than one, it is not strictly a proportion, but it does tend to increase with *ab*.

To summarize the R^2 indices suggested by MacKinnon (2008) and Lindenberger and Pötter (1998), none can be interpreted as proportions. On the other hand, $R^2_{4.6}$ does fall between 0 and 1 inclusively, and its magnitude does correspond to that of *ab* (the relationship is slightly concave up). All of the indices suggested by MacKinnon (2008) are standardized and amenable to confidence interval construction.

Despite the obvious appeal of R^2 indices as effect size indices, Fichman (1999) reviews several reasons why researchers may wish to be cautious when using R^2 indices to compare theories. According to Fichman (1999), R^2 indices are not always useful for comparing rival theories, can easily be misapplied or used inconsistently, leading to overinterpretations or underinterpretations of effect size, are context-dependent (Balluerka et al., 2005), and are often less intuitive and more difficult to evaluate than one might think. Researchers often focus on explained variance, but in so doing they often neglect to understand the underlying process itself. Furthermore, explained variance depends on how much variance there is to explain (Fern & Monroe, 1996; Henson, 2006; Nakagawa & Cuthill, 2007), and this quantity may differ between studies, between populations, and between manipulated versus observed versions of the same variable, precluding the use of R^2 indices for meaningfully comparing effects. Ozer (1985) cautioned that R^2 may not be interpretable as a proportion of variance in many circumstances, which undermines any effect size index that depends on this interpretation. Further, Sechrest and Yeaton (1982) pointed out that researchers often assume that the amount of variance to be explained is 100%. However, this assumption is rarely met in practice because few variables are measured without error. The explainable variance in *Y* is often much less than 100%. Sechrest and Yeaton (1982) also pointed out that it is often difficult to decide on the appropriate effect size to use, and different treatment strengths can result in very different effect sizes.

Finally, it could be argued that because population, rather than sample, effect sizes are the true quantities of interest, then the researcher ought to adjust these R^2 indices for positive bias (resulting from using sample values to estimate population quantities) if they are to be used at all. For example, Ezekiel (1930) described an adjusted R^2 index $\tilde{R}^2_{Y.\mathbf{X}} = 1-(1-R^2_{Y.\mathbf{X}})\left(\frac{n-1}{n-m}\right)$, where *n* is the sample size, *m* is the number of regression parameters (intercept

[5] In order for $R^2_{4.6}$ to exactly equal 1, *X* and *M* would have to be perfectly correlated, and the squared semipartial correlation of *Y* with *M* would have to be exactly 1. Because this cannot occur without introducing perfect collinearity, 1 is a limiting value and is not actually obtainable in practice.

[6] Tatsuoka (1973, p. 281) reminded us that "the product of two proportions is itself a meaningful proportion only when the second proportion is based on that subset of the universe that is 'earmarked' by the first proportion."

[7] Matrices were generated using a fast Markov chain neighborhood sampling method that retains generated matrices meeting a positive minimum eigenvalue criterion. For more information, see Preacher (2006). We selected 15,000 matrices to visually convey the relative density of points in different regions of the plots.

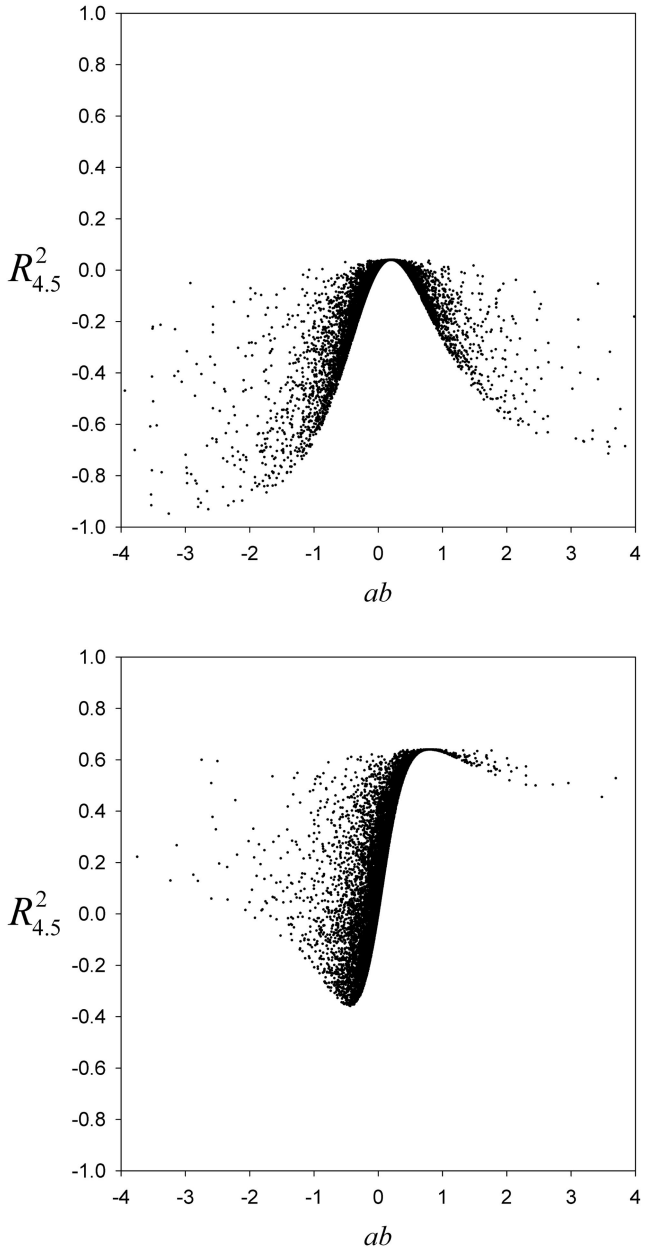

Figure 4. Plots of $R^2_{4.5}$ plotted against *ab* for 15,000 indirect effects, holding the total effect *c* constant at .2 (top) and .8 (bottom).

and slopes), and **X** is a vector of regressors. The formula for $R^2_{4.5}$ incorporating these adjustments would thus be

$$\tilde{R}^2_{4.5} = \left(1 - (1 - r^2_{YM})\frac{n-1}{n-2}\right) - \left(\left(1 - (1 - R^2_{Y,MX})\frac{n-1}{n-3}\right) - \left(1 - (1 - r^2_{YX})\frac{n-1}{n-2}\right)\right)$$

$$= 1 + (1 - R^2_{Y,MX})\frac{n-1}{n-3} + (r^2_{YX} + r^2_{YM} - 2)\frac{n-1}{n-2}. \quad (19)$$

Owing to the moderately large sample size of $n = 432$ in the SPBY data, $\tilde{R}^2_{4.5} = .0333$—not very different from the unadjusted value of $R^2_{4.5} = .0338$. In smaller samples, such adjustments would be more noticeable. Bias adjusted versions of $R^2_{4.6}$ and $R^2_{4.7}$ are

$$\tilde{R}^2_{4.6} = \left(1 - (1 - r^2_{YM})\frac{n-1}{n-2}\right)\left(1 - \left(\frac{1 - R^2_{Y,MX}}{1 - r^2_{YX}}\right)\frac{n-2}{n-3}\right), \quad (20)$$

and

$$\tilde{R}^2_{4.7} = \frac{\left(1 - (1 - r^2_{YM})\frac{n-1}{n-2}\right)\left(1 - \left(\frac{1 - R^2_{Y,MX}}{1 - r^2_{YX}}\right)\frac{n-2}{n-3}\right)}{\left(1 - (1 - R^2_{Y,MX})\frac{n-1}{n-3}\right)}, \quad (21)$$

respectively. See Wang and Thompson (2007) for an extended discussion of Ezekiel's (1930) and other potential adjustments to r^2 and R^2.

Hansen and McNeal's (1996) Effect Size Index for Two Groups

Many applications of mediation analysis involve a binary *X* (such as gender or experimental condition), where the purpose of the analysis is to determine whether and to what extent the mean

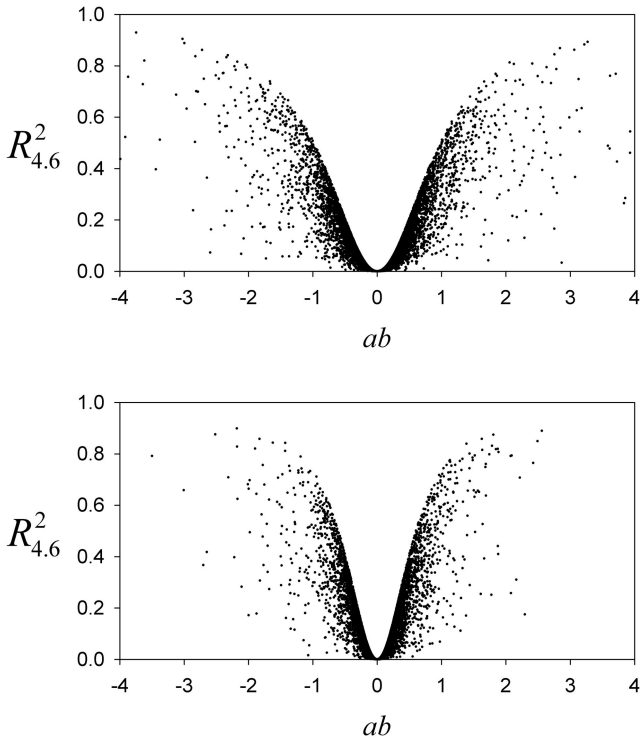

Figure 5. Plots of $R^2_{4.6}$ plotted against *ab* for 15,000 indirect effects, holding the total effect *c* constant at .2 (top) and .8 (bottom).

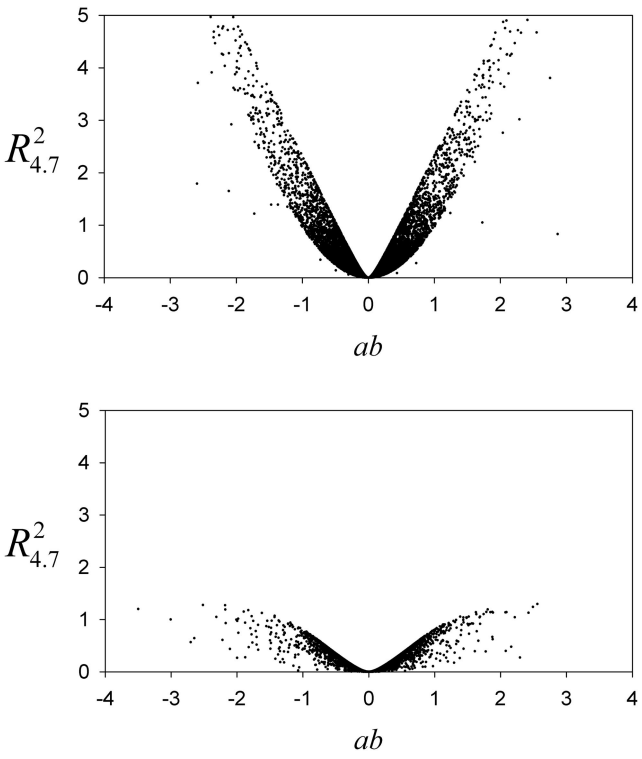

Figure 6. Plots of $R^2_{4.7}$ plotted against *ab* for 15,000 indirect effects, holding the total effect *c* constant at .2 (top) and .8 (bottom).

difference in *Y* can be attributed to *X* indirectly through a mediator *M*. Hansen and McNeal (1996) suggested an effect size index for mediation that can be obtained by applying a sample size adjustment to Sobel's (1982) test statistic in such two-group designs. When *X* is a binary variable,

$$ES = \frac{ab}{\sqrt{a^2 s_b^2 + b^2 s_a^2}} \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}, \quad (22)$$

where n_1 and n_2 are the sample sizes of Group 1 and Group 2, respectively, and s_a and s_b are the standard errors of the regression coefficients *a* and *b*, respectively. Sample values are substituted for their population counterparts. Note that sample size is introduced in the denominator of Sobel's statistic by including the s^2 terms. The intent of the multiplier added by Hansen and McNeal is to remove that influence of sample size, rendering an index that does not depend on *n*. *ES* (effect size) is, in fact, relatively robust to large shifts in sample size. However, use of the *ES* index is limited to settings in which *X* is binary. In addition, because the statistic is not bounded, standardized, or robust to changes in scale, it is unclear how to interpret it.

New Methods of Expressing Effect Size for Mediation Effects

Two alternative approaches avoid some of the problems inherent in informal descriptors and ratio measures. These effect sizes conform more closely to the definition and desiderata of good effect size measures identified earlier than do the measures described in the previous section.

A Residual-Based Index

The first new effect size we consider elaborates on a method proposed by Berry and Mielke (2002) for effect size computation in univariate or multivariate regression models. Their original method involves computing functions of residuals for models conforming to a null and alternative hypothesis, obtaining their ratio, and subtracting the result from 1. We propose an index that combines information about the variance in *M* explained by *X* and the variance in *Y* explained by both *X* and *M*.

Berry and Mielke consider regression models conforming to null and alternative hypotheses. In the univariate case where *M* is regressed on a number of *X* variables, the null and alternative models are, respectively (for case *i*'s data),

$$M_i = \sum_{j=1}^{m_0} X_{ij}\beta_{0j} + e_{0i} \text{ and } M_i = \sum_{j=1}^{m_1} X_{ij}\beta_{1j} + e_{1i}, \quad (23, 24)$$

where m_0 is the number of regressors under the null hypothesis, m_1 is the number of regressors under the alternative hypothesis ($m_1 > m_0$), *i* indexes cases, β_{0j} and β_{1j} are coefficients for the X_{ij} regressors in the null and alternative models, respectively, and all variables are mean-centered so that intercepts can be omitted. Residuals for the null and alternative models are given by

$$e_{0i} = M_i - \sum_{j=1}^{m_0} X_{ij}\beta_{0j} \text{ and } e_{1i} = M_i - \sum_{j=1}^{m_1} X_{ij}\beta_{1j}, \quad (25, 26)$$

respectively. The effect size is then computed as $1 - \sum_{i=1}^{n}\sqrt{e_{1i}^2} / \sum_{i=1}^{n}\sqrt{e_{0i}^2} = 1 - \sum_{i=1}^{n}|e_{1i}| / \sum_{i=1}^{n}|e_{0i}|$. Because the denominator sum will always exceed the numerator sum, Berry and Mielke's (2002) effect size necessarily lies between 0 and 1.

Mediation analysis, on the other hand, involves residuals for the *M* equation and the *Y* equation. Researchers often expect that *X* will explain a large amount of variance in both *M* and *Y* and that *M* will explain the same variance in *Y* that *X* explains. Therefore, the null scenario in mediation analysis is one in which there is no explanation of variance in *M* or *Y*. The limiting alternative scenario, on the other hand, is one in which *X* explains all of the variance in *M*, while *X* and *M* each explain all of the variance in *Y*. The observed effect size will lie between these two extremes (0 and 1). These extreme values suggest a basis for defining the residuals to be used in a modification of Berry and Mielke's (2002) index appropriate for mediation analysis.

First, we define the null model residuals for the *M* and *Y* equations (in which no variance is explained in either) as

$$e_{0M_i} = M_i - \overline{M} \quad (27)$$

and

$$e_{0Y_i} = Y_i - \overline{Y}, \tag{28}$$

respectively, where $\overline{M}$ and $\overline{Y}$ are the means of M and Y. Second, we define alternative model residuals for the M and Y equations (conforming to the estimated model) as

$$\begin{aligned} e_{1M_i} &= e_{M.X_i} \\ &= M_i - d_{M.X} - aX_i \end{aligned} \tag{29}$$

and

$$\begin{aligned} e_{1Y_i} &= e_{Y.X_i} + e_{Y.M_i} - e_{Y.XM_i} \\ &= (Y_i - d_{Y.X} - cX_i) + (Y_i - d_{Y.M} - dM_i) - (Y_i - d_{Y.XM} \\ &\qquad - bM_i - c'X_i) \\ &= Y_i - d_{Y.X} - cX_i - d_{Y.M} - dM_i + d_{Y.XM} + bM_i + c'X_i, \end{aligned} \tag{30}$$

respectively, where a, b, and c' are as defined earlier and d is the slope relating M to Y with no other regressors in the model. The residuals e_{1Y_i} correspond to that part of Y not explained jointly by X and M. Therefore, e_{1Y_i} is the part of Y not explained by X, plus the part of Y not explained by M, minus the part these two quantities share (so that it is not counted twice). Equation 30 is analogous to the way in which a joint probability is determined, where two probabilities are added and their intersection removed [i.e., P(A or B) = P(A) + P(B) – P(A and B)]. Ideally, the e_{1Y_i}s will be as small as possible. These residuals are then combined to produce Γ, a residual-based effect size index:

$$\Gamma = 1 - \frac{\sum_{i=1}^{n}\left(\sqrt{e_{1M_i}^2} + \sqrt{e_{1Y_i}^2}\right)}{\sum_{i=1}^{n}\left(\sqrt{e_{0M_i}^2} + \sqrt{e_{0Y_i}^2}\right)} = \frac{\sum_{i=1}^{n}(|e_{1M_i}| + |e_{1Y_i}|)}{\sum_{i=1}^{n}(|e_{0M_i}| + |e_{0Y_i}|)} \tag{31}$$

Γ can be interpreted as a measure of the extent to which variance in M is explained by X, and variance in Y is explained jointly by X and M. It has the advantages of being directly interpretable and lying on a meaningfully scaled metric; Γ is bounded above by 1 and is very rarely less than 0 when mediation is in evidence. G, the sample estimate of Γ, is also independent of sample size. Whereas confidence intervals may be constructed for Γ using bootstrap methods, as of yet, no exact analytic confidence interval formulation procedure is known to us. In the SPBY example, $G = \hat{\Gamma} =$ .049 (95% CI [.024, .081]).

One complicating factor should be noted with respect to G: The value of G is influenced by the scales of M and Y. If these scales differ, then G will be unduly influenced by either the residuals associated with M or those associated with Y. Therefore, we suggest a standardized version, γ (g in samples), that has the same formula but draws residuals from standardized regressions rather than unstandardized regressions (i.e., replaces the errors in Equation 31 with those obtained from using standardized scores instead of raw scores in the regression model). That is, γ (or g) is Equation 31 applied to the residuals of regression models in which all of the variables have been standardized. In the SPBY example, $g = \hat{\gamma} =$ .044 (95% CI [.023, .072]).

A second complicating factor associated with Γ and γ is that they can be nonzero in situations where the indirect effect is absent (i.e., $ab = 0$ but Γ and γ are nonzero). Nevertheless, we do not consider nonzero residual-based effect sizes (Γ or γ) necessarily problematic. If one considers the theoretically ideal mediation effect as one in which X explains all the variance in M and both X and M explain all the variance in Y, then it is sensible to quantify how close to that ideal we have come. The effect sizes Γ and γ quantify this idea. This is one case in which the effect size measure does not coincide with the way in which the effect itself is commonly operationalized—it is a measure of total variance explained rather than a product of regression coefficients. Therefore, we suggest that Γ and γ can serve as useful supplementary measures to report along with the indirect effect and other effect sizes, such as the unstandardized and standardized maximum possible indirect effect, which we now discuss.

Maximum Possible Indirect Effect and Its Standardized Version

The second effect size we propose, and ultimately recommend, is the magnitude of the indirect effect relative to the maximum possible indirect effect. In general, an effect that may seem trivial in absolute size may in fact be relatively large when one considers the range of potential values the effect *could have* assumed, given characteristics of the design or distributional characteristics of the variables. Even under ideal distributional conditions and linear relationships, there are real limits on the values that regression weights (and thus indirect effects) can take on, given certain characteristics of the data.

For example, consider a multiple regression model that accounts for "only" .125 (raw) units of variance in the dependent variable. Initially, accounting for only .125 units of variance may seem trivial. However, if the variance of the dependent variable were only .15 units to begin with, the model accounts for 83.33% (.125/.15 = .8333) of the variance that it *could have* possibly accounted for. Thus, looking at the raw value of the amount of variance accounted for does not necessarily give an accurate portrayal of the effectiveness of a regressor.

As another example, this time in the context of mediation, consider the hypothetical situation in which $s_X^2 = s_M^2 = s_Y^2 = 1.0$ and the total effect $c = .6$. Given these constraints, ab is not bounded because b is not bounded. However, consider the case in which we hold a fixed to some conditional value, like .3. When this is true, b is bounded (in fact, b must lie between $\pm.84$), and therefore ab is also bounded (here, to $\pm.25$). Similarly, for a given value of b under the above constraints, the absolute value of a must lie within a certain range, and therefore ab is again bounded. The range of possible standardized indirect effects is presented graphically on the vertical axis of Figure 7 for $c = -.19$ (the standardized c coefficient from the SPBY example). From Figure 7 it can be seen that in the neighborhood of $a = 0$, the possible range of ab is restricted to the neighborhood of $ab = 0$. As a departs from 0 in

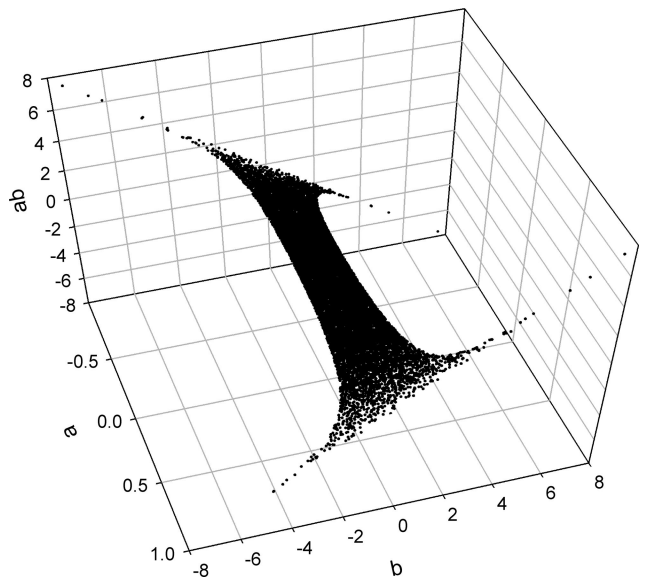

Figure 7. Plot of the indirect effect *ab* versus *a* and *b* when *X*, *M*, and *Y* are standardized and $c = -0.19$.

either direction, larger values of *b* become possible, in turn permitting a greater potential range for values of *ab*.

A logical question, then, is how can these bounds on *a*, *b*, and *ab* be determined? Hubert (1972) demonstrated how to obtain lower and upper boundaries for elements of a covariance matrix. Consider the 3×3 symmetric matrix $\mathbf{S}$ (which in the present case may be considered the covariance matrix of *X*, *M*, and *Y*), partitioned as

$$\mathbf{S} = \begin{bmatrix} \mathbf{A} & \mathbf{G} \\ \mathbf{G}' & \text{var}(Y) \end{bmatrix} = \left[\begin{array}{cc|c} \sigma_X^2 & \sigma_{MX} & \sigma_{YX} \\ \sigma_{MX} & \sigma_M^2 & \sigma_{YM} \\ \hline \sigma_{YX} & \sigma_{YM} & \sigma_Y^2 \end{array}\right]. \quad (32)$$

$\mathbf{S}$ is nonnegative definite if and only if $\mathbf{G}'\mathbf{A}^{-1}\mathbf{G} \leq \text{var}(Y)$. This restriction implies the following permissible range for the *a* coefficient of a mediation model if *b* and *c* are held fixed:

$$a \in \left\{ \frac{\sigma_{YM}\sigma_{YX} \pm \sqrt{\sigma_M^2\sigma_Y^2 - \sigma_{YM}^2}\sqrt{\sigma_X^2\sigma_Y^2 - \sigma_{YX}^2}}{\sigma_X^2\sigma_Y^2} \right\} \quad (33)$$

(where $\in$ here means "is contained in") and the following permissible range for the *b* coefficient if *a* and *c* are held fixed:

$$b \in \left\{ \pm \frac{\sqrt{\sigma_X^2\sigma_Y^2 - \sigma_{YX}^2}}{\sqrt{\sigma_X^2\sigma_M^2 - \sigma_{MX}^2}} \right\}. \quad (34)$$

Given these restrictions, it is possible to derive boundaries for the indirect effect *ab* given a fixed *a* and *c* or a fixed *b* and *c*. First, let $\mathfrak{M}(\cdot)$ be an operator that returns the most extreme possible observable value of the argument parameter with the same sign as the corresponding sample parameter estimate. For example, if $\hat{b} = -.10$ and the bounds identified for *b* in Equation 34 are $-.21$ and .21, $\mathfrak{M}(b) = -.21$. $\mathfrak{M}(b)$ would not be .21 because $\hat{b}$ is negative, necessitating that $\mathfrak{M}(b)$ also be negative. Holding *b* and *c* constant, the bounds on *ab* can be derived by beginning with the bounds implied for *a* and multiplying by the $\mathfrak{M}(b)$ identified in Equation 34 by the most extreme possible value with the same sign as *ab*. This yields (after a few algebraic steps)

$$ab \in \left\{ \mathfrak{M}(b) \frac{\sigma_{YM}\sigma_{YX} \pm \sqrt{\sigma_M^2\sigma_Y^2 - \sigma_{YM}^2}\sqrt{\sigma_X^2\sigma_Y^2 - \sigma_{YX}^2}}{\sigma_X^2\sigma_Y^2} \right\}. \quad (35)$$

Taking the most extreme limit of the two limits from Equation 35 that is of the same sign as *ab* provides the maximum possible indirect effect. Holding *a* and *c* constant, the equivalent bounds on *ab* can be derived by beginning with the bounds implied for *b* and multiplying by $\mathfrak{M}(a)$ obtained from Equation 33, yielding

$$ab \in \left\{ \pm \mathfrak{M}(a) \frac{\sqrt{\sigma_X^2\sigma_Y^2 - \sigma_{YX}^2}}{\sqrt{\sigma_X^2\sigma_M^2 - \sigma_{MX}^2}} \right\}. \quad (36)$$

As above, taking the most extreme of the two limits from Equation 36 that is of the same sign as *ab* provides the maximum possible indirect effect. Rather than determining the possible range of *ab*, the maximum possible indirect effect is obtained by the product of $\mathfrak{M}(a)$ and $\mathfrak{M}(b)$:

$$\mathfrak{M}(ab) = \mathfrak{M}(a)\mathfrak{M}(b). \quad (37)$$

Full derivations of these results can be found in Appendix A. The obtained indirect effect $\hat{a}\hat{b}$ can be interpreted in light of this range. $\mathfrak{M}(ab)$ will be identical for both of these methods. Notice also that $\mathfrak{M}(ab)$ can itself be used as an effect size, even though we primarily suggest that it be used as the standardizer in the calculation of another effect size we present below.

In sum, if the research question involves the effect size of an indirect effect, it is sensible to ask what the maximum attainable value of the indirect effect (in the direction of the observed indirect effect) *could have been*, conditional on the sample variances and on the magnitudes of relationships among some of the variables.[8] Reporting that an indirect effect is $ab = .57$ tells us little in isolation (much like the amount of variance accounted for in the previous example), but when it is considered that the most extreme value *ab* could possibly have attained (given the observed *c* and conditioning on either *a* or *b*) is .62, the effect size may be considered larger than if $\mathfrak{M}(ab)$ were .86.[9]

As an example of computing $\mathfrak{M}(ab)$ in the SPBY data, first note that the covariance matrix of VAC, ATD, and DVB is

$$\mathbf{S} = \begin{bmatrix} 2.2683 & 0.6615 & -0.0869 \\ 0.6615 & 2.2764 & -0.2259 \\ -0.0869 & -0.2259 & 0.0922 \end{bmatrix}. \quad (38)$$

The permissible ranges of *a* and *b* are thus

[8] In addition to restrictions imposed by the magnitudes of certain variances and coefficients, there is a further restriction on the possible size of an indirect effect. Carroll (1961) and Breaugh (2003) pointed out that, unless the two variables have equivalent distributions (e.g., both normal), their correlation cannot equal 1.0. Because variables are rarely perfectly normally (or even equally) distributed in real applications, the maximum possible effect usually will be lower in practice than in theory.

[9] If only *c* is held to be known (rather than either {*a* and *c*} or {*b* and *c*}), these results imply a bounded region for *ab*.

$$a \in \left\{ \frac{s_{YM}s_{YX} \pm \sqrt{s_M^2 s_Y^2 - s_{YM}^2}\sqrt{s_X^2 s_Y^2 - s_{YX}^2}}{s_X^2 s_Y^2} \right\}$$

$$\in \left\{ \frac{\left[\begin{array}{r} -.2259 \times -.0869 \pm \sqrt{(2.2764)(.0922) - (-.2259)^2} \\ \times \sqrt{(2.2683)(.0922) - (-.0869)^2} \end{array} \right]}{(2.2683)(.0922)} \right\}$$

$$\in \{-.762, .950\}, \quad (39)$$

making $\mathfrak{M}(a) = .950$, and

$$b \in \left\{ \pm \frac{\sqrt{s_X^2 s_Y^2 - s_{YX}^2}}{\sqrt{s_X^2 s_M^2 - s_{MX}^2}} \right\}$$

$$\in \left\{ \pm \frac{\sqrt{(2.2683)(.0922) - (-.0869)^2}}{\sqrt{(2.2683)(2.2764) - (.6615)^2}} \right\}$$

$$\in \{-.207, .207\}, \quad (40)$$

making $\mathfrak{M}(b) = -.207$. The sample bounds for ab are obtained using Equation 35 and the outer bound for b:

$$ab \in \left\{ \mathfrak{M}(b) \left(\frac{s_{YM}s_{YX} \pm \sqrt{s_M^2 s_Y^2 - s_{YM}^2}\sqrt{s_X^2 s_Y^2 - s_{YX}^2}}{s_X^2 s_Y^2} \right) \right\}$$

$$\in \{-.2065(-.7618, .9495)\}$$

$$\in \{-.196, .157\}, \quad (41)$$

making $\mathfrak{M}(ab) = -.196$. Instead, using Equation 36 and the outer bound for a, the sample bounds are

$$ab \in \left\{ \pm \mathfrak{M}(a) \frac{\sqrt{\sigma_X^2 \sigma_Y^2 - \sigma_{YX}^2}}{\sqrt{\sigma_X^2 \sigma_M^2 - \sigma_{MX}^2}} \right\}$$

$$\in \left\{ \pm .9495 \frac{\sqrt{(2.2683)(.0922) - (-.0869)^2}}{\sqrt{(2.2683)(2.2764) - (.6615)^2}} \right\}$$

$$\in \{-.196, .196\}, \quad (42)$$

making $\mathfrak{M}(ab) = -.196$, which was already known from Equation 41. Regardless of whether $\mathfrak{M}(ab)$ is calculated directly based on Equation 37 or indirectly based on Equation 35 or Equation 36, the value will always be the same.

Given that $\mathfrak{M}(ab) = -.196$, the observed $\hat{a}\hat{b}$ of $-.028$ implies that even though the indirect effect is statistically significant, it is much smaller than it *could have* been. This is a key point: Bounding values of parameters often are not appreciated when interpreting the magnitude and importance of effect sizes.

Rather than considering the maximum value of the indirect effect as an effect size, per se, we use $\mathfrak{M}(ab)$ to define a standardized effect size that compares the value of ab to $\mathfrak{M}(ab)$. That is, we define the standardized effect size, which we denote κ^2,

$$\kappa^2 = \frac{ab}{\mathfrak{M}(ab)}. \quad (43)$$

κ^2 is interpreted as *the proportion of the maximum possible indirect effect* that could have occurred, had the constituent effects been as large as the design and data permitted. $\kappa^2 = 0$ implies that there is no linear indirect effect, and $\kappa^2 = 1$ implies that the indirect effect is as large as it potentially could have been. We use the notation kappa-squared (i.e., κ^2) to denote that like the squared multiple correlation coefficient, it (a) cannot be negative, (b) is bounded (inclusively) between 0 and 1, and (c) represents the proportion of the value of a quantity to the maximum value it could have been. Otherwise, κ^2 and the population squared multiple correlation coefficient have generally different properties. In order to estimate κ^2, we suggest that sample values of the variances and covariances replace their population counterparts. κ^2 is a standardized value, as it is not wedded to the original scale of the variables, allows (at least) bootstrap confidence intervals to be formed, and is independent of sample size. We find these qualities to be advantageous. For the SPBY example, the proportion of the maximum observed indirect effect that was observed is

$$k^2 = \hat{\kappa}^2 = \frac{\hat{a}\hat{b}}{\mathfrak{M}(\hat{a}\hat{b})} = \frac{-.0281}{-.1961} = .143, \quad (44)$$

with bootstrap 95% CI [.100, .190].

R Tools

To encourage and facilitate the application of the methods we have advocated for communicating the effect size of mediation effects, we have developed a set of easy to use R functions, which are contained in the MBESS (Kelley & Lai, 2010; Kelley, 2007a, 2007b) R (R Development Core Team, 2010) package. The specific MBESS functions are `mediation()`, `mediation.effect.bar.plot()`, and `mediation.effect.plot()`, which implement the mediation model and all of the mediation effect sizes we have discussed, with or without bootstrap confidence intervals. The functions `mediation.effect.bar.plot()` and `mediation.effect.plot()` can be used to create effect bar plots and effect plots, respectively—two graphical methods of communicating mediation effects (discussed on the website). The `mediation()` function accepts either raw data or summary statistics (i.e., means and variances/covariances) for simple mediation models, as we have described. The `mediation()` function reports the results of the three separate regression models and all of the effect sizes, optionally with percentile and/or bias corrected accelerated bootstrap confidence intervals. Documentation for the functions is contained within the MBESS package.

Discussion

Researchers should consider not only the statistical significance of indirect effects but also the effect size of a given effect. We reemphasize the growing consensus that reporting effect size is crucial to the advancement of psychological science. As Cumming et al. (2007) wrote,

> It is important and urgent that psychology change its emphasis from the dichotomous decision making of NHST to estimation of effect size . . . Effect sizes must always be reported—in an appropriate measure, and wherever possible with CIs—and then interpreted. To achieve this goal,

researchers need further detailed guidance, examples of good practice, and editorial or institutional leadership. (pp. 231–232)

It is hoped that this discussion has been a step in the right direction in the context of reporting and interpreting mediation effects. This is an especially important type of statistical model in which to apply effect sizes, as mediation models are so widely used in research.

We have discussed many effect sizes with potential application in mediation analysis. The researcher may be at somewhat of a loss when choosing an appropriate effect size measure, given that there are so many choices. We offer two suggestions that may render the choice easier. First, there is no reason to report only one effect size. If circumstances permit, reporting multiple effect sizes can yield greater understanding of a given effect, with the added benefit that more effect size measures are available for possible use in meta-analysis. As an analogy, regression results are often reported in a table containing unstandardized regression coefficients, standardized regression coefficients, and ΔR^2 for each regressor, R^2, and $R^2_{Adj.}$—five different types of effect sizes, with the first three effect size measures being repeated for each regressor in the model. Each of these effect sizes measures communicates different information in different units. Additionally, a researcher desiring to communicate the meaning of an indirect effect in a mediation analysis might also report the unstandardized indirect effect (*ab*) and the residual-based index γ, or some other combination of effect sizes.

Second, earlier we presented three desiderata for good effect size measurement: A good effect size should be scaled on a meaningful, but not necessarily standardized, metric; it should be amenable to the construction of confidence intervals; and it should be independent, or nearly so, of sample size. The researcher should remain cognizant of these desiderata when selecting an appropriate effect size. We suggest that if the researcher wishes to use an effect size that does not fulfill all the desiderata we have outlined, it should be supplemented with additional effect sizes.

We encourage researchers to think about the most important aspects of the effects they wish to report and seek effect size measures that address those aspects. To aid researchers in deciding which effect size measure(s) to report, in Table 3 we note, for each effect size measure, whether it fulfills the three desiderata. More concretely, we recommend researchers report, at a minimum, the estimated value of κ^2, the ratio of the obtained indirect effect to the maximum possible indirect effect. The benefits of using κ^2 are that it is standardized, in the sense that its value is not wedded to the particular scale used in the mediation analysis; it is on an interpretable metric (0 to 1); it is insensitive to sample size; and with bootstrap methods, it allows for the construction of confidence intervals. We do not rule out an analytic method of confidence interval formation for κ^2, but for practical purposes, the bootstrap confidence interval is advantageous.

An obvious question regarding κ^2 is "what constitutes a large value?" As we have previously noted, a "large" value need not constitute an important value, and an important value need not be a "large" value. We also are very hesitant to put any qualitative descriptors on a quantitative value. However, if one were forced to attach such labels to κ^2, we believe it makes sense to interpret them in the same light as squared correlation coefficients are often interpreted, that is, with Cohen's (1988) guidelines. In particular, after some hesitation on the part of Cohen to define benchmarks for various effect sizes (1988, section 1.4), he ultimately concludes that such benchmarks can be beneficial. For the proportion of variance accounted for in one variable by another (i.e., r^2_{xy}), Cohen defines small, medium, and large effect sizes as .01, .09, and .25

Table 3
Characteristics of 16 Effect Size Measures for Mediation Analysis

Effect size	Standardized?	Bounded?	Desideratum 1: Interpretable scaling?	Desideratum 2: Confidence interval available?	Desideratum 3: Independent of sample size?
Verbal descriptors					
P_M	✓			✓	✓
R_M	✓			✓	✓
S_M				✓	✓
ab			✓	✓	✓
ab_{ps}	Partially		✓	✓	✓
ab_{cs}	✓		✓	✓	✓
$R^2_{4.5}$	✓			✓	✓
$R^2_{4.6}$	✓	✓		✓	✓
$R^2_{4.7}$	✓			✓	✓
SOS	✓			✓	✓
ES	✓			✓	✓
Γ		Partially	✓	✓	✓
γ	✓	Partially	✓	✓	✓
$\mathfrak{M}(ab)$			✓	✓	✓
κ^2	✓	✓	✓	✓	✓

Note. *SOS* = shared over simple effects; *ES* = effect size.

(pp. 79–81). Because of the similar properties of r^2_{xy} and κ^2, we believe that the benchmarks for r^2_{xy} are similarly applicable for κ^2. Recalling that in the SPBY data $\kappa^2 = .143$ with 95% CI [.100, .190], one could argue that the mediation effect in the SPBY data is at least medium (because the 95% confidence interval excludes .09) but smaller than large (because the confidence interval excludes .25). Thus, the size of the mediation effect in the SPBY data may be appropriately labeled as lying in the medium range. However, we emphasize that the best way to describe κ^2 is with its quantitative value, estimated to be .143 for the SPBY data.

To truly understand the value of κ^2 in a given context, comprehensive studies describing the typical values of κ^2 in well-defined research areas would be very useful. Further, such effect sizes could be treated as dependent variables with various regressors/explanatory variables in a meta-analytic context, where an explanation of various values of effect sizes is attempted.

Limitations and Cautions

It is appropriate at this point to identify several limitations and cautions in the application of effect sizes for mediation effects. First, as is the case with virtually any effect size, relatively small effect sizes may be substantively important, whereas relatively large ones may be trivial, depending on the research context. An objectively small effect in high-stakes research may be deemed very important by the scientific community, whereas an objectively large effect in other fields may not reach a noteworthy level. Because of this, we caution researchers to not rigidly interpret effect size measures against arbitrary benchmarks. Snyder and Lawson (1993) emphasized that using benchmarks to judge effect size estimates ignores judgments regarding clinical significance, the researcher's personal value system, the research questions posed, societal concerns, and the design of a particular study. Although we do not argue against setting benchmarks, it is important that the field of application to which the benchmarks apply should be clearly delineated. Further, a strong rationale should be given for why a particular value is given for a benchmark. Probably the safest route is to simply report the effect size without providing unnecessary and possibly misleading commentary about its size (Robinson, Whittaker, Williams, & Beretvas, 2003).

Second, we caution that it is a mistake to equate effect size with practical importance or clinical significance (Thompson, 2002). Certainly some values of some effect sizes can convey practical importance, but depending on the particular situation, what is and what is not practically important will vary. Fern and Monroe (1996, pp. 103–104) cautioned that importance or substantive significance should not be inferred solely on the basis of the magnitude of an effect size. Several features of the research context should be considered as well. Ultimately, the practical importance of an effect depends on the research context, the cost of data collection, the importance of the outcome variable, and the likely impact of the results. Consequently, researchers are cautioned to avoid generalizing beyond the particular research design employed. Effect sizes should serve only as guides to practical importance, not as replacements for it, and are at best imperfect shadows of the true practical importance of an effect.

Third, outliers and violations of assumptions of statistical methods compromise effect size estimates, *p*-values, and confidence intervals. Correspondingly, it is vitally important that researchers perform diagnostic checks to ensure that the assumptions of their inferential techniques are not obviously violated. It is well known that outliers can spuriously inflate or deflate statistical significance, Type II error rates (Wilcox, 2005), and confidence interval coverages, but they can also inflate or deflate estimates of effect size. Consequently, it is wise to determine the extent to which assumptions are met and to examine one's data for outliers. If problems are detected, remedial steps should be taken, or appropriate caveats should be included with the reported results.

Fourth, all of the effect sizes we have discussed have limitations. It is important to keep those limitations in mind when using them. For example, the effect sizes discussed have not yet been extended for use in models involving multiple mediators. No effect size is universally applicable or meaningful in all contexts. Correspondingly, researchers will need to decide which effect size most appropriately conveys the meaning of the results in the particular context.

Fifth, effect sizes can depend on variability. Brandstätter (1999) pointed out that the "degree of manipulation" can affect the value of the effect size. Cortina and Dunlap (1997) and Dooling and Danks (1975) made similar points. This realization is important for effect sizes in the context of mediation because *X* frequently is manipulated, yet the strength of the manipulation often is made arbitrarily large to maximize power for detecting an effect. The effect size for such effects does not imply that a "large" effect would be similarly astounding had *X* merely been observed rather than manipulated. In fact, McClelland (1997) and McClelland and Judd (1993) advocated an "extreme groups" approach for *detecting* effects, such that extremes are oversampled at the expense of central scores. Oversampling extreme groups is a worthwhile approach when the goal is to maximize power in order to infer that differences exist. However, trustworthy and generalizable estimates of standardized effect size require (a) random sampling or (b) manipulation strength that matches what one would expect to find in nature (see Cortina & DeShon, 1998, for a summary of some of these points).

Future Directions

The methods we have discussed here are hardly definitive. For example, results here are limited to the simple mediation model. Extension to more complex mediation models, such as those for panel data (Cole & Maxwell, 2003), moderated indirect effects (Edwards & Lambert, 2007; Preacher, Rucker, & Hayes, 2007), or multiple mediators (MacKinnon, 2000; Preacher & Hayes, 2008b) should be devised and investigated. As Maxwell and Cole (2007) pointed out, $\hat{P}_M$ is a biased estimate of effect size if one uses cross-sectional data when the effect of interest is one that takes time to unfold. They go on to show that when *X* has greater longitudinal stability than *M*, $\hat{P}_M$ will be biased downward relative to the corresponding longitudinal index. Conversely, when *M* is more stable than *X*, $\hat{P}_M$ will be biased upward. This criticism is valid, and similar criticisms apply to any effect size measure based on the analysis of cross-sectional data when the process under study is a longitudinal one. The lesson here is that any effect size estimate must be interpreted in the context of the specific research design used. The specific lags chosen to separate the measurement

of *X*, *M*, and *Y* are part of that context, so generalizing results beyond that context should be done with extreme caution.

Particularly useful would be studies conducted to establish defensible benchmarks for different effect size measures denoting small, medium, and large effects in particular research contexts. For example, a study could be conducted to establish what values of ab_{cs} or κ^2 should be considered small, medium, and large for alcoholism treatment studies to help determine what mechanisms are primarily responsible for explaining the effectiveness of intervention programs. The establishment of generally accepted benchmarks based on published research for different effect sizes in a variety of research contexts would facilitate meta-analysis. We believe κ^2, in addition to other effect sizes, will be a useful measure in meta-analyses of mediation effects when the proportion of the maximum possible indirect effect obtainable across different samples is an interesting research question. κ^2 fulfills the desiderata for good effect size estimates, and it is standardized (and therefore independent of the scaling of variables) and bounded.

We have not discussed sample size planning methods for mediation models, but it is an important issue. The power analytic (e.g., Cohen, 1988) and the accuracy in parameter estimation (AIPE) approaches to sample size planning (e.g., Kelley & Maxwell, 2003, 2008) should be considered. Theoretically, for any effect of interest, sample size can be planned so that there is a sufficiently high probability to reject a false null hypothesis (i.e., power analysis) and/or sample size can be planned so that the confidence interval is sufficiently narrow (i.e., accuracy in parameter estimation; see Maxwell, Kelley, & Rausch, 2008, for a review). "Whenever an estimate is of interest, so too should the corresponding confidence interval for the population quantity" (Kelley, 2008, p. 553). The goal of AIPE is to obtain a sufficiently narrow confidence interval that conveys the accuracy with which the population value has been estimated by the point estimate of the effect size. If that confidence interval is wide for an effect size of interest, less is known about the value of the population parameter than would be desirable. Moving forward, the power and AIPE approaches to sample size planning should be fully developed for effect sizes used in a mediation context.

Prescriptions for Research

Research on effect size for mediation effects is relatively new and thus not fully developed. We nevertheless end by offering some concrete recommendations for researchers wanting to report effect size for mediation effects. We reiterate the "Three Reporting Rules" suggested by Vacha-Haase and Thompson (2004) for reporting effect size estimates; these rules are just as applicable in the mediation context as in many other contexts:

1. Be explicit about what effect size is being reported. Often we see "effect size" reported with no indication as to whether the reported index is a correlation coefficient, mean difference, Cohen's *d*, η^2, and so on, or why one measure was chosen over competing measures. The particular effect size cannot always be accurately inferred from the context in which it was reported.

2. Interpret effect sizes considering both their assumptions and limitations. All of the effect sizes discussed here require certain assumptions to be satisfied in order to obtain trustworthy confidence intervals. Specifically, observations should be independent and identically distributed, or the researcher risks obtaining confidence intervals with incorrect coverage. In addition, we have discussed limitations associated with each of the effect sizes we presented. It is important to explicitly consider the assumptions and limitations when reporting and interpreting effect size.

3. Report confidence intervals for population effect sizes. Confidence intervals are necessary to communicate the degree of sampling uncertainty associated with estimates of effect size and are a valuable adjunct to any point estimate effect size measure.

Our most emphatic recommendation, however, is that methodologists undertake more research to establish meaningful, trustworthy methods of communicating effect size and practical importance for mediation effects. Tests of mediation have proliferated at an unprecedented rate in recent years, with a heavy emphasis on establishing statistical significance and very little attention devoted to quantifying effect size and/or practical importance. We fear that this lack of balance has led to a proliferation of nonchance but trivially important mediation effects being reported in the literature. In addition, the lack of effect size reporting for mediation analyses has seriously limited the accumulation of knowledge in some fields. Consequently, we strongly urge researchers to consider not only whether their effects are due to chance (i.e., is statistical significance reached?) but also how large the effect sizes are and how relevant they are to theory or practice.

References

Abelson, R. P. (1995). *Statistics as principled argument.* Hillsdale, NJ: Erlbaum.

Achen, C. H. (1977). Measuring representation: Perils of the correlation coefficient. *American Journal of Political Science, 21,* 805–815. doi: 10.2307/2110737

Albert, J. M. (2008). Mediation analysis via potential outcomes models. *Statistics in Medicine, 27,* 1282–1304. doi:10.1002/sim.3016

Alwin, D. F., & Hauser, R. M. (1975). The decomposition of effects in path analysis. *American Sociological Review, 40,* 37–47. doi:10.2307/2094445

American Educational Research Association. (2006). *Standards for reporting on empirical social science research in AERA publications.* Washington, DC: Author.

American Psychological Association. (2010). *Publication manual of the American Psychological Association* (6th ed.). Washington, DC: Author.

Baguley, T. (2009). Standardized or simple effect size: What should be reported? *British Journal of Psychology, 100,* 603–617. doi:10.1348/000712608X377117

Balluerka, N., Gómez, J., & Hidalgo, D. (2005). The controversy over null hypothesis significance testing revisited. *Methodology, 1,* 55–70.

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. doi:10.1037/0022-3514.51.6.1173

Barreto, M., & Ellemers, N. (2005). The burden of benevolent sexism: How it contributes to the maintenance of gender inequalities. *European Journal of Social Psychology, 35,* 633–642. doi:10.1002/ejsp.270

Berry, K. J., & Mielke, P. W., Jr. (2002). Least sum of Euclidean regression residuals: Estimation of effect size. *Psychological Reports, 91,* 955–962. doi:10.2466/PR0.91.7.955-962

Bird, K. D. (2002). Confidence intervals for effect sizes in analysis of variance. *Educational and Psychological Measurement, 62,* 197–226. doi:10.1177/0013164402062002001

Bobko, P., & Rieck, A. (1980). Large sample estimators for standard errors of functions of correlation coefficients. *Applied Psychological Measurement, 4,* 385–398. doi:10.1177/014662168000400309

Bond, C. F., Jr., Wiitala, W. L., & Richard, F. D. (2003). Meta-analysis of raw mean differences. *Psychological Methods, 8,* 406–418. doi:10.1037/1082-989X.8.4.406

Brandstätter, E. (1999). Confidence intervals as an alternative to significance testing. *Methods of Psychological Research Online, 4,* 33–46.

Breaugh, J. A. (2003). Effect size estimation: Factors to consider and mistakes to avoid. *Journal of Management, 29,* 79–97. doi:10.1177/014920630302900106

Buyse, M., & Molenberghs, G. (1998). Criteria for the validation of surrogate endpoints in randomized experiments. *Biometrics, 54,* 1014–1029. doi:10.2307/2533853

Carroll, J. B. (1961). The nature of data, or how to choose a correlation coefficient. *Psychometrika, 26,* 347–372. doi:10.1007/BF02289768

Cheung, M. W.-L. (2009). Comparison of methods for constructing confidence intervals of standardized indirect effects. *Behavior Research Methods, 41,* 425–438. doi:10.3758/BRM.41.2.425

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). New York, NY: Academic Press.

Cole, D. A., & Maxwell, S. E. (2003). Testing mediational models with longitudinal data: Questions and tips in the use of structural equation modeling. *Journal of Abnormal Psychology, 112,* 558–577. doi:10.1037/0021-843X.112.4.558

Cortina, J. M., & DeShon, R. P. (1998). Determining relative importance of predictors with the observational design. *Journal of Applied Psychology, 83,* 798–804. doi:10.1037/0021-9010.83.5.798

Cortina, J. M., & Dunlap, W. P. (1997). On the logic and purpose of significance testing. *Psychological Methods, 2,* 161–172. doi:10.1037/1082-989X.2.2.161

Cumming, G., Fidler, F., Leonard, M., Kalinowski, P., Christiansen, A., Kleinig, A., . . . Wilson, S. (2007). Statistical reform in psychology: Is anything changing? *Psychological Science, 18,* 230–232. doi:10.1111/j.1467-9280.2007.01881.x

Cumming, G., & Finch, S. (2001). A primer on the understanding, use and calculation of confidence intervals that are based on central and noncentral distributions. *Educational and Psychological Measurement, 61,* 532–574.

Ditlevsen, S., Christensen, U., Lynch, J., Damsgaard, M. T., & Keiding, N. (2005). The mediation proportion: A structural equation approach for estimating the proportion of exposure effect on outcome explained by an intermediate variable. *Epidemiology, 16,* 114–120. doi:10.1097/01.ede.0000147107.76079.07

Dooling, D., & Danks, J. H. (1975). Going beyond tests of significance: Is psychology ready? *Bulletin of the Psychonomic Society, 5,* 15–17.

Edwards, J. R., & Lambert, L. S. (2007). Methods for integrating moderation and mediation: A general path analytic framework using moderated path analysis. *Psychological Methods, 12,* 1–22. doi:10.1037/1082-989X.12.1.1

Efron, B. (1987). Better bootstrap confidence intervals. *Journal of the American Statistical Association, 82,* 171–185. doi:10.2307/2289144

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

Ezekiel, M. (1930). *Methods of correlation analysis.* New York, NY: Wiley.

Fairchild, A. J., MacKinnon, D. P., Taborga, M. P., & Taylor, A. B. (2009). R^2 effect-size measures for mediation analysis. *Behavior Research Methods, 41,* 486–498. doi:10.3758/BRM.41.2.486

Fern, E. F., & Monroe, K. B. (1996). Effect-size estimates: Issues and problems in interpretation. *The Journal of Consumer Research, 23,* 89–105. doi:10.1086/209469

Fichman, M. (1999). Variance explained: Why size does not (always) matter. *Research in Organizational Behavior, 21,* 295–331.

Fidler, F., & Thompson, B. (2001). Computing correct confidence intervals for ANOVA fixed- and random-effects effect sizes. *Educational and Psychological Measurement, 61,* 575–604.

Freedman, L. S. (2001). Confidence intervals and statistical power of the "Validation" ratio for surrogate or intermediate endpoints. *Journal of Statistical Planning and Inference, 96,* 143–153. doi:10.1016/S0378-3758(00)00330-X

Frick, R. W. (1999). Defending the statistical status quo. *Theory & Psychology, 9,* 183–189. doi:10.1177/095935439992002

Greenland, S. (1998). Meta-analysis. In K. J. Rothman & S. Greenland (Eds.), *Modern epidemiology* (pp. 643–674). Philadelphia, PA: Lippincott-Raven.

Greenland, S., Schlesselman, J. J., & Criqui, M. H. (1986). The fallacy of employing standardized regression coefficients and correlations as measures of effect. *American Journal of Epidemiology, 123,* 203–208.

Grissom, R. J., & Kim, J. J. (2005). *Effect size for research: A broad practical approach.* Mahwah, NJ: Erlbaum.

Hansen, W. B., & McNeal, R. B. (1996). The law of maximum expected potential effect: Constraints placed on program effectiveness by mediator relationships. *Health Education Research, 11,* 501–507. doi:10.1093/her/11.4.501

Henson, R. K. (2006). Effect-size measures and meta-analytic thinking in counseling psychology research. *The Counseling Psychologist, 34,* 601–629. doi:10.1177/0011000005283558

Huang, B., Sivaganesan, S., Succop, P., & Goodman, E. (2004). Statistical assessment of mediational effects for logistic mediational models. *Statistics in Medicine, 23,* 2713–2728. doi:10.1002/sim.1847

Hubert, L. J. (1972). A note on the restriction of range for Pearson product-moment correlation coefficients. *Educational and Psychological Measurement, 32,* 767–770. doi:10.1177/001316447203200315

James, L. R., & Brett, J. M. (1984). Mediators, moderators, and tests for mediation. *Journal of Applied Psychology, 69,* 307–321. doi:10.1037/0021-9010.69.2.307

Jessor, R., & Jessor, S. L. (1977). *Problem behavior and psychosocial development: A longitudinal study of youth.* New York, NY: Academic Press.

Jessor, R., & Jessor, S. L. (1991). *Socialization of problem behavior in youth* [Data file and code book]. Henry A. Murray Research Archive (http://www.murray.harvard.edu/), Cambridge, MA.

Kelley, K. (2007a). Confidence intervals for standardized effect sizes: Theory, application, and implementation. *Journal of Statistical Software, 20,* 1–24.

Kelley, K. (2007b). Methods for the behavioral, educational, and social sciences: An R package. *Behavior Research Methods, 39,* 979–984.

Kelley, K. (2008). Sample size planning for the squared multiple correlation coefficient: Accuracy in parameter estimation via narrow confidence intervals. *Multivariate Behavioral Research, 43,* 524–555. doi:10.1080/00273170802490632

Kelley, K., & Lai, K. (2010). MBESS (Version 3.2.0) [Computer software and manual]. Retrieved from http://www.cran.r-project.org/

Kelley, K., & Maxwell, S. E. (2003). Sample size for multiple regression: Obtaining regression coefficients that are accurate, not simply significant. *Psychological Methods, 8,* 305–321. doi:10.1037/1082-989X.8.3.305

Kelley, K., & Maxwell, S. E. (2008). Power and accuracy for omnibus and targeted effects: Issues of sample size planning with applications to multiple regression. In P. Alasuuta, L. Bickman, & J. Brannen (Eds.), *The Sage handbook of social research methods* (pp. 166–192). Newbury Park, CA: Sage.

Kim, J.-O., & Ferree, G. D., Jr. (1981). Standardization in causal analysis. *Sociological Methods & Research, 10,* 187–210.

King, G. (1986). How not to lie with statistics: Avoiding common mistakes in quantitative political science. *American Journal of Political Science, 30,* 666–687. doi:10.2307/2111095

Kirk, R. E. (1996). Practical significance: A concept whose time has come. *Educational and Psychological Measurement, 56,* 746–759. doi:10.1177/0013164496056005002

Lin, D. Y., Fleming, T. R., & De Gruttola, V. (1997). Estimating the proportion of treatment effect explained by a surrogate marker. *Statistics in Medicine, 16,* 1515–1527. doi:10.1002/(SICI)1097-0258(19970715)16:13::AID-SIM5723.0.CO;2-1

Lindenberger, U., & Pötter, U. (1998). The complex nature of unique and shared effects in hierarchical linear regression: Implications for developmental psychology. *Psychological Methods, 3,* 218–230. doi:10.1037/1082-989X.3.2.218

MacKinnon, D. P. (1994). Analysis of mediating variables in prevention intervention studies. In A. Cazares & L. A. Beatty (Eds.), *Scientific methods for prevention intervention research* (pp. 127–153; DHHS Publication 94–3631). *NIDA Research Monograph, 139*.

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* (pp. 141–160). Mahwah, NJ: Erlbaum.

MacKinnon, D. P. (2008). *Introduction to statistical mediation analysis.* Mahwah, NJ: Erlbaum.

MacKinnon, D. P., & Dwyer, J. H. (1993). Estimating mediated effects in prevention studies. *Evaluation Review, 17,* 144–158. doi:10.1177/0193841X9301700202

MacKinnon, D. P., Fairchild, A. J., & Fritz, M. S. (2007). Mediation analysis. *Annual Review of Psychology, 58,* 593–614. doi:10.1146/annurev.psych.58.110405.085542

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. doi:10.1037/1082-989X.7.1.83

MacKinnon, D. P., Warsi, G., & Dwyer, J. H. (1995). A simulation study of mediated effect measures. *Multivariate Behavioral Research, 30,* 41–62. doi:10.1207/s15327906mbr3001_3

Mathieu, J. E., & Taylor, S. R. (2006). Clarifying conditions and decision points for mediational type inferences in organizational behavior. *Journal of Organizational Behavior, 27,* 1031–1056. doi:10.1002/job.406

Maxwell, S. E., & Cole, D. A. (2007). Bias in cross-sectional analyses of longitudinal mediation. *Psychological Methods, 12,* 23–44. doi:10.1037/1082-989X.12.1.23

Maxwell, S. E., Kelley, K., & Rausch, J. R. (2008). Sample size planning for statistical power and accuracy in parameter estimation. *Annual Review of Psychology, 59,* 537–563. doi:10.1146/annurev.psych.59.103006.093735

McClelland, G. H. (1997). Optimal design in psychological research. *Psychological Methods, 2,* 3–19. doi:10.1037/1082-989X.2.1.3

McClelland, G. H., & Judd, C. M. (1993). Statistical difficulties of detecting interactions and moderator effects. *Psychological Bulletin, 114,* 376–390. doi:10.1037/0033-2909.114.2.376

Moher, D., Hopewell, S., Schulz, K., Montori, V., Gøtzsche, P. C., Devereaux, P. J., . . . Altman, M. G. (2010). Consort 2010 explanation and elaboration: Updated guidelines for reporting parallel group randomized trials. *British Medical Journal, 340,* 698–702. doi:10.1136/bmj.c869

Nakagawa, S., & Cuthill, I. C. (2007). Effect size, confidence interval and statistical significance: A practical guide for biologists. *Biological Reviews, 82,* 591–605. doi:10.1111/j.1469-185X.2007.00027.x

National Center for Education Statistics. (2003). *NCES statistical standards.* Washington, DC: Department of Education.

O'Grady, K. E. (1982). Measures of explained variance: Cautions and limitations. *Psychological Bulletin, 92,* 766–777. doi:10.1037/0033-2909.92.3.766

Ozer, D. (1985). Correlation and the coefficient of determination. *Psychological Bulletin, 97,* 307–315. doi:10.1037/0033-2909.97.2.307

Ozer, D. J. (2007). Evaluating effect size in personality research. In R. W. Robins, R. C. Fraley, & R. F. Krueger (Eds.), *Handbook of research methods in personality psychology* (pp. 495–501). New York, NY: The Guilford Press.

Preacher, K. J. (2006). Quantifying parsimony in structural equation modeling. *Multivariate Behavioral Research, 41,* 227–259. doi:10.1207/s15327906mbr4103_1

Preacher, K. J., & Hayes, A. F. (2008a). Contemporary approaches to assessing mediation in communication research. In A. F. Hayes, M. D. Slater, & L. B. Snyder (Eds.), *The Sage sourcebook of advanced data analysis methods for communication research* (pp. 13–54). Thousand Oaks, CA: Sage.

Preacher, K. J., & Hayes, A. F. (2008b). Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behavior Research Methods, 40,* 879–891. doi:10.3758/BRM.40.3.879

Preacher, K. J., Rucker, D. D., & Hayes, A. F. (2007). Assessing moderated mediation hypotheses: Theory, methods, and prescriptions. *Multivariate Behavioral Research, 42,* 185–227.

Raykov, T., Brennan, M., Reinhardt, J. P., & Horowitz, A. (2008). Comparison of mediated effects: A correlation structure modeling approach. *Structural Equation Modeling, 15,* 603–626. doi:10.1080/10705510802339015

R Development Core Team. (2010). *R: A language and environment for statistical computing.* Vienna, Austria: Author.

Robinson, D. H., Whittaker, T. A., Williams, N. J., & Beretvas, S. N. (2003). It's not effect sizes so much as comments about their magnitude that mislead readers. *The Journal of Experimental Education, 72,* 51–64. doi:10.1080/00220970309600879

Sechrest, L., & Yeaton, W. H. (1982). Magnitudes of experimental effects in social science research. *Evaluation Review, 6,* 579–600. doi:10.1177/0193841X8200600501

Shrout, P. E., & Bolger, N. (2002). Mediation in experimental and nonexperimental studies: New procedures and recommendations. *Psychological Methods, 7,* 422–445. doi:10.1037/1082-989X.7.4.422

Smithson, M. (2001). Correct confidence intervals for various regression effect sizes and parameters: The importance of noncentral distributions in computing intervals. *Educational and Psychological Measurement, 61,* 605–632. doi:10.1177/00131640121971392

Snyder, P., & Lawson, S. (1993). Evaluating results using corrected and uncorrected effect size estimates. *Journal of Experimental Education, 61,* 334–349.

Sobel, M. E. (1982). Asymptotic confidence intervals for indirect effects in structural equation models. In S. Leinhardt (Ed.), *Sociological methodology* 1982 (pp. 290–312). Washington, DC: American Sociological Association.

Tatsuoka, M. M. (1973). Multivariate analysis in education research. *Review of Research in Education, 1,* 273–319.

Thompson, B. (2002). What future quantitative social science research could look like: Confidence intervals for effect sizes. *Educational Researcher, 31,* 25–32. doi:10.3102/0013189X031003025

Thompson, B. (2007). Effect sizes, confidence intervals, and confidence intervals for effect sizes. *Psychology in the Schools, 44,* 423–432. doi:10.1002/pits.20234

Tofighi, D., MacKinnon, D. P., & Yoon, M. (2009). Covariances between regression coefficient estimates in a single mediator model. *British Journal of Mathematical and Statistical Psychology, 62,* 457–484. doi:10.1348/000711008X331024

Vacha-Haase, T., Nilsson, J. E., Reetz, D. R., Lance, T. S., & Thompson, B. (2000). Reporting practices and APA editorial policies regarding

statistical significance and effect size. *Theory & Psychology, 10,* 413–425. doi:10.1177/0959354300103006

Vacha-Haase, T., & Thompson, B. (2004). How to estimate and interpret various effect sizes. *Journal of Counseling Psychology, 51,* 473–481. doi:10.1037/0022-0167.51.4.473

Wang, Y., & Taylor, J. M. G. (2002). A measure of the proportion of treatment effect explained by a surrogate marker. *Biometrics, 58,* 803–812. doi:10.1111/j.0006-341X.2002.00803.x

Wang, Z., & Thompson, B. (2007). Is the Pearson r^2 biased, and if so, what is the best correction formula? *The Journal of Experimental Education, 75,* 109–125. doi:10.3200/JEXE.75.2.109-125

Wilcox, R. (2005). *Introduction to robust estimation and hypothesis testing* (2nd ed.). San Diego, CA: Academic Press.

Wilkinson, L., & the Task Force on Statistical Inference. (1999). Statistical methods in psychology journals. *American Psychologist, 54,* 594–604. doi:10.1037/0003-066X.54.8.594

Appendix A

Derivation of Boundaries for Maximum Possible Indirect Effect

Correlations within a correlation matrix set limits on the ranges of the remaining correlations because of the necessity to maintain positive definiteness. These range restrictions, in turn, imply range restrictions on unstandardized regression weights subject to the variables' variances. Beginning with correlations in a 3×3 matrix,

$$\rho_{21}\rho_{32} - \sqrt{1-\rho_{21}^2}\sqrt{1-\rho_{32}^2} \leq \rho_{31} \leq \rho_{21}\rho_{32} + \sqrt{1-\rho_{21}^2}\sqrt{1-\rho_{32}^2}, \tag{A1}$$

$$\rho_{31}\rho_{32} - \sqrt{1-\rho_{31}^2}\sqrt{1-\rho_{32}^2} \leq \rho_{21} \leq \rho_{31}\rho_{32} + \sqrt{1-\rho_{31}^2}\sqrt{1-\rho_{32}^2}, \tag{A2}$$

$$\rho_{21}\rho_{31} - \sqrt{1-\rho_{21}^2}\sqrt{1-\rho_{31}^2} \leq \rho_{32} \leq \rho_{21}\rho_{31} + \sqrt{1-\rho_{21}^2}\sqrt{1-\rho_{31}^2}. \tag{A3}$$

For the simple mediation model considered in this article, in which X, M, and Y are variables 1, 2, and 3, the corresponding standardized regression weights are

$$a = \rho_{21}, \tag{A4}$$

$$b = \frac{\rho_{32} - \rho_{21}\rho_{31}}{1-\rho_{21}^2}, \tag{A5}$$

$$c' = \frac{\rho_{31} - \rho_{21}\rho_{32}}{1-\rho_{21}^2}. \tag{A6}$$

The unstandardized regression weights are

$$a = \rho_{21}\frac{\sigma_M}{\sigma_X}, \tag{A7}$$

$$b = \frac{\rho_{32} - \rho_{21}\rho_{31}}{1-\rho_{21}^2}\frac{\sigma_Y}{\sigma_M}, \tag{A8}$$

$$c' = \frac{\rho_{31} - \rho_{21}\rho_{32}}{1-\rho_{21}^2}\frac{\sigma_Y}{\sigma_X}. \tag{A9}$$

The unstandardized indirect effect is therefore

$$\begin{aligned} ab &= \rho_{21}\frac{\rho_{32} - \rho_{21}\rho_{31}}{1-\rho_{21}^2}\frac{\sigma_Y}{\sigma_X} \\ &= \frac{\sigma_{XM}(\sigma_X^2\sigma_{MY} - \sigma_{XM}\sigma_{XY})}{\sigma_M^2(\sigma_X^2)^2(1-\sigma_{XM}^2)}. \end{aligned} \tag{A10}$$

(*Appendix continues*)

Now, consider the partitioned matrix:

$$\mathbf{\Sigma} = \begin{bmatrix} \mathbf{A} & \mathbf{G} \\ \mathbf{G}' & \text{var}(Y) \end{bmatrix}. \tag{A11}$$

$\mathbf{\Sigma}$ is nonnegative definite if and only if $\mathbf{G}'\mathbf{A}^{-1}\mathbf{G} \leq \text{var}(Y)$. Hubert (1972) showed the special case where $\mathbf{\Sigma}=\mathbf{P}$, a correlation matrix:

$$\mathbf{P} = \begin{bmatrix} \mathbf{A} & \mathbf{G} \\ \mathbf{G}' & \text{var}(Y) \end{bmatrix} = \left[\begin{array}{cc|c} 1 & \rho_{21} & \rho_{31} \\ \rho_{21} & 1 & \rho_{32} \\ \hline \rho_{31} & \rho_{32} & 1 \end{array}\right]. \tag{A12}$$

In this special case, the theorem implies

$$\frac{1}{1-\rho_{21}^2}\begin{bmatrix}\rho_{31} & \rho_{32}\end{bmatrix}\begin{bmatrix} 1 & -\rho_{21} \\ -\rho_{21} & 1 \end{bmatrix}\begin{bmatrix}\rho_{31} \\ \rho_{32}\end{bmatrix} \leq 1,$$

$$\frac{1}{1-\rho_{21}^2}\begin{bmatrix}\rho_{31}-\rho_{21}\rho_{32} & \rho_{32}-\rho_{21}\rho_{31}\end{bmatrix}\begin{bmatrix}\rho_{31} \\ \rho_{32}\end{bmatrix} \leq 1,$$

$$\frac{\rho_{31}(\rho_{31}-\rho_{21}\rho_{32})+\rho_{32}(\rho_{32}-\rho_{21}\rho_{31})}{1-\rho_{21}^2} \leq 1,$$

$$\rho_{21}^2+\rho_{31}^2+\rho_{32}^2-2\rho_{21}\rho_{31}\rho_{32}-1 \leq 0, \tag{A13}$$

which can be solved algebraically (by completing the square) to obtain any of the three ranges from above (Equations A1, A2, and A3). In the more general case of $\mathbf{\Sigma}$, we can obtain bounds for, say, σ_{MX}:

$$\Sigma = \left[\begin{array}{cc} \mathbf{A} & \mathbf{G} \\ \mathbf{G}' & \text{var}(Y) \end{array}\right] = \left[\begin{array}{cc|c} \sigma_X^2 & \sigma_{MX} & \sigma_{YX} \\ \sigma_{MX} & \sigma_M^2 & \sigma_{YM} \\ \hline \sigma_{YX} & \sigma_{YM} & \sigma_Y^2 \end{array}\right], \tag{A14}$$

implying

$$\frac{1}{\sigma_X^2\sigma_M^2-\sigma_{MX}^2}\begin{bmatrix}\sigma_{YX} & \sigma_{YM}\end{bmatrix}\begin{bmatrix} \sigma_M^2 & -\sigma_{MX} \\ -\sigma_{MX} & \sigma_X^2 \end{bmatrix}\begin{bmatrix}\sigma_{YX} \\ \sigma_{YM}\end{bmatrix} \leq \sigma_Y^2,$$

$$\frac{1}{\sigma_X^2\sigma_M^2-\sigma_{MX}^2}\begin{bmatrix}\sigma_{YX}\sigma_M^2-\sigma_{MX}\sigma_{YM} & -\sigma_{MX}\sigma_{YX}+\sigma_{YM}\sigma_X^2\end{bmatrix}\begin{bmatrix}\sigma_{YX} \\ \sigma_{YM}\end{bmatrix} \leq \sigma_Y^2,$$

$$\sigma_{MX}^2\sigma_Y^2-2\sigma_{MX}\sigma_{YM}\sigma_{YX} \leq \sigma_X^2\sigma_M^2\sigma_Y^2-\sigma_{YM}^2\sigma_X^2-\sigma_{YX}^2\sigma_M^2,$$

$$\sigma_{MX}^2-\frac{2\sigma_{MX}\sigma_{YM}\sigma_{YX}}{\sigma_Y^2} \leq \frac{\sigma_X^2\sigma_M^2\sigma_Y^2-\sigma_{YM}^2\sigma_X^2-\sigma_{YX}^2\sigma_M^2}{\sigma_Y^2},$$

$$\sigma_{MX}^2-\frac{2\sigma_{MX}\sigma_{YM}\sigma_{YX}}{\sigma_Y^2}+\left(\frac{\sigma_{YM}\sigma_{YX}}{\sigma_Y^2}\right)^2 \leq \frac{\sigma_X^2\sigma_M^2\sigma_Y^2-\sigma_{YM}^2\sigma_X^2-\sigma_{YX}^2\sigma_M^2}{\sigma_Y^2}+\left(\frac{\sigma_{YM}\sigma_{YX}}{\sigma_Y^2}\right)^2,$$

$$\left(\sigma_{MX}-\frac{\sigma_{YM}\sigma_{YX}}{\sigma_Y^2}\right)^2 \leq \frac{\sigma_X^2\sigma_M^2\sigma_Y^2-\sigma_{YM}^2\sigma_X^2-\sigma_{YX}^2\sigma_M^2}{\sigma_Y^2}+\left(\frac{\sigma_{YM}\sigma_{YX}}{\sigma_Y^2}\right)^2,$$

$$\sigma_{MX} \in \left\{\frac{\sigma_{YM}\sigma_{YX}}{\sigma_Y^2} \pm \sqrt{\frac{\sigma_X^2\sigma_M^2\sigma_Y^2-\sigma_{YM}^2\sigma_X^2-\sigma_{YX}^2\sigma_M^2}{\sigma_Y^2}+\left(\frac{\sigma_{YM}\sigma_{YX}}{\sigma_Y^2}\right)^2}\right\},$$

$$\sigma_{MX} \in \left\{\frac{\sigma_{YM}\sigma_{YX} \pm \sqrt{\sigma_M^2\sigma_Y^2-\sigma_{YM}^2}\sqrt{\sigma_X^2\sigma_Y^2-\sigma_{YX}^2}}{\sigma_Y^2}\right\}, \tag{A15}$$

(*Appendix continues*)

with $\in$ meaning "is contained in." Ranges for the other two covariances are of similar form. The correlation case is a special case of this more general treatment for covariances.

The bounds implied for regression coefficient a can be derived from the above result by simply isolating a using its expression in covariance metric:

$$\sigma_{MX} \in \left\{ \frac{\sigma_{YM}\sigma_{YX} \pm \sqrt{\sigma_M^2\sigma_Y^2 - \sigma_{YM}^2}\sqrt{\sigma_X^2\sigma_Y^2 - \sigma_{YX}^2}}{\sigma_Y^2} \right\},$$

$$\frac{\sigma_{MX}}{\sigma_X^2} \in \left\{ \frac{\sigma_{YM}\sigma_{YX} \pm \sqrt{\sigma_M^2\sigma_Y^2 - \sigma_{YM}^2}\sqrt{\sigma_X^2\sigma_Y^2 - \sigma_{YX}^2}}{\sigma_X^2\sigma_Y^2} \right\},$$

$$a \in \left\{ \frac{\sigma_{YM}\sigma_{YX} \pm \sqrt{\sigma_M^2\sigma_Y^2 - \sigma_{YM}^2}\sqrt{\sigma_X^2\sigma_Y^2 - \sigma_{YX}^2}}{\sigma_X^2\sigma_Y^2} \right\}. \tag{A16}$$

Another method for obtaining the bounds for a, using its correlation metric expression and altering the central term until it equals the formula for a and simplifying, is

$$\rho_{31}\rho_{32} - \sqrt{1-\rho_{31}^2}\sqrt{1-\rho_{32}^2} \le \rho_{21} \le \rho_{31}\rho_{32} + \sqrt{1-\rho_{31}^2}\sqrt{1-\rho_{32}^2},$$

$$\rho_{31}\rho_{32}\frac{\sigma_M}{\sigma_X} - \sqrt{1-\rho_{31}^2}\sqrt{1-\rho_{32}^2}\frac{\sigma_M}{\sigma_X} \le \rho_{21}\frac{\sigma_M}{\sigma_X} \le \rho_{31}\rho_{32}\frac{\sigma_M}{\sigma_X} + \sqrt{1-\rho_{31}^2}\sqrt{1-\rho_{32}^2}\frac{\sigma_M}{\sigma_X},$$

$$\frac{\sigma_{YX}\sigma_{YM}}{\sigma_X^2\sigma_Y^2} - \sqrt{1-\left(\frac{\sigma_{YX}}{\sigma_X\sigma_Y}\right)^2}\sqrt{1-\left(\frac{\sigma_{YM}}{\sigma_M\sigma_Y}\right)^2}\frac{\sigma_M\sigma_X\sigma_Y^2}{\sigma_X^2\sigma_Y^2} \le a \le \frac{\sigma_{YX}\sigma_{YM}}{\sigma_X^2\sigma_Y^2} + \sqrt{1-\left(\frac{\sigma_{YX}}{\sigma_X\sigma_Y}\right)^2}\sqrt{1-\left(\frac{\sigma_{YM}}{\sigma_M\sigma_Y}\right)^2}\frac{\sigma_M\sigma_X\sigma_Y^2}{\sigma_X^2\sigma_Y^2},$$

$$\frac{\sigma_{YX}\sigma_{YM}}{\sigma_X^2\sigma_Y^2} - \sqrt{1-\frac{\sigma_{YX}^2}{\sigma_X^2\sigma_Y^2}}\sqrt{1-\frac{\sigma_{YM}^2}{\sigma_M^2\sigma_Y^2}}\frac{\sigma_M\sigma_X\sigma_Y^2}{\sigma_X^2\sigma_Y^2} \le a \le \frac{\sigma_{YX}\sigma_{YM}}{\sigma_X^2\sigma_Y^2} + \sqrt{1-\frac{\sigma_{YX}^2}{\sigma_X^2\sigma_Y^2}}\sqrt{1-\frac{\sigma_{YM}^2}{\sigma_M^2\sigma_Y^2}}\frac{\sigma_M\sigma_X\sigma_Y^2}{\sigma_X^2\sigma_Y^2},$$

$$\frac{\sigma_{YX}\sigma_{YM} - \sqrt{\sigma_Y^2\sigma_X^2 - \sigma_{YX}^2}\sqrt{\sigma_M^2\sigma_Y^2 - \sigma_{YM}^2}}{\sigma_X^2\sigma_Y^2} \le a \le \frac{\sigma_{YX}\sigma_{YM} + \sqrt{\sigma_Y^2\sigma_X^2 - \sigma_{YX}^2}\sqrt{\sigma_M^2\sigma_Y^2 - \sigma_{YM}^2}}{\sigma_X^2\sigma_Y^2},$$

$$a \in \left\{ \frac{\sigma_{YX}\sigma_{YM} \pm \sqrt{\sigma_Y^2\sigma_X^2 - \sigma_{YX}^2}\sqrt{\sigma_M^2\sigma_Y^2 - \sigma_{YM}^2}}{\sigma_X^2\sigma_Y^2} \right\}. \tag{A17}$$

For b, a similar procedure could be followed:

$$\rho_{21}\rho_{31} - \sqrt{1-\rho_{21}^2}\sqrt{1-\rho_{31}^2} \le \rho_{32} \le \rho_{21}\rho_{31} + \sqrt{1-\rho_{21}^2}\sqrt{1-\rho_{31}^2},$$

$$-\sqrt{1-\rho_{21}^2}\sqrt{1-\rho_{31}^2} \le \rho_{32} - \rho_{21}\rho_{31} \le \sqrt{1-\rho_{21}^2}\sqrt{1-\rho_{31}^2},$$

$$-\frac{\sqrt{1-\rho_{21}^2}\sqrt{1-\rho_{31}^2}}{1-\rho_{21}^2} \le \frac{\rho_{32} - \rho_{21}\rho_{31}}{1-\rho_{21}^2} \le \frac{\sqrt{1-\rho_{21}^2}\sqrt{1-\rho_{31}^2}}{1-\rho_{21}^2},$$

$$-\frac{\sigma_Y\sqrt{1-\rho_{31}^2}}{\sigma_M\sqrt{1-\rho_{21}^2}} \le b \le \frac{\sigma_Y\sqrt{1-\rho_{31}^2}}{\sigma_M\sqrt{1-\rho_{21}^2}}, \tag{A18}$$

(*Appendix continues*)

$$b \in \left\{ \pm \frac{\sigma_Y \sqrt{1 - \frac{\sigma_{YX}^2}{\sigma_Y^2 \sigma_X^2}}}{\sigma_M \sqrt{1 - \frac{\sigma_{MX}^2}{\sigma_M^2 \sigma_X^2}}} \right\},$$

$$b \in \left\{ \pm \frac{\sqrt{\sigma_X^2 \sigma_Y^2 - \sigma_{YX}^2}}{\sqrt{\sigma_X^2 \sigma_M^2 - \sigma_{MX}^2}} \right\}.$$

Now that bounds are known for b (given a and c) and for a (given b and c), the bounds for ab can be determined. For given a and c, the bounds on ab can be derived by beginning with the bounds implied for b and multiplying all terms by the conditional value $\mathfrak{M}(a)$, the most extreme possible observable value of a with the same sign as $\hat{a}$ (from Equation A16 or A17):

$$-\frac{\sqrt{\sigma_X^2 \sigma_Y^2 - \sigma_{YX}^2}}{\sqrt{\sigma_X^2 \sigma_M^2 - \sigma_{MX}^2}} < b < \frac{\sqrt{\sigma_X^2 \sigma_Y^2 - \sigma_{YX}^2}}{\sqrt{\sigma_X^2 \sigma_M^2 - \sigma_{MX}^2}},$$

$$-\mathfrak{M}(a)\frac{\sqrt{\sigma_X^2 \sigma_Y^2 - \sigma_{YX}^2}}{\sqrt{\sigma_X^2 \sigma_M^2 - \sigma_{MX}^2}} < b\,\mathfrak{M}(a) < \mathfrak{M}(a)\frac{\sqrt{\sigma_X^2 \sigma_Y^2 - \sigma_{YX}^2}}{\sqrt{\sigma_X^2 \sigma_M^2 - \sigma_{MX}^2}},$$

$$ab \in \left\{ \pm \mathfrak{M}(\mathrm{a}) \frac{\sqrt{\sigma_X^2 \sigma_Y^2 - \sigma_{YX}^2}}{\sqrt{\sigma_X^2 \sigma_M^2 - \sigma_{MX}^2}} \right\}. \tag{A19}$$

For given b and c, the bounds on ab can be derived by beginning with the bounds implied for a and multiplying all terms by the conditional value $\mathfrak{M}(b)$:

$$\frac{\sigma_{YX}\sigma_{YM} - \sqrt{\sigma_Y^2 \sigma_X^2 - \sigma_{YX}^2}\sqrt{\sigma_M^2 \sigma_Y^2 - \sigma_{YM}^2}}{\sigma_X^2 \sigma_Y^2} \leq a \leq \frac{\sigma_{YX}\sigma_{YM} + \sqrt{\sigma_Y^2 \sigma_X^2 - \sigma_{YX}^2}\sqrt{\sigma_M^2 \sigma_Y^2 - \sigma_{YM}^2}}{\sigma_X^2 \sigma_Y^2},$$

$$\mathfrak{M}(b)\frac{\sigma_{YX}\sigma_{YM} - \sqrt{\sigma_Y^2 \sigma_X^2 - \sigma_{YX}^2}\sqrt{\sigma_M^2 \sigma_Y^2 - \sigma_{YM}^2}}{\sigma_X^2 \sigma_Y^2} \leq a\,\mathfrak{M}(b) \leq \mathfrak{M}(b)\frac{\sigma_{YX}\sigma_{YM} + \sqrt{\sigma_Y^2 \sigma_X^2 - \sigma_{YX}^2}\sqrt{\sigma_M^2 \sigma_Y^2 - \sigma_{YM}^2}}{\sigma_X^2 \sigma_Y^2},$$

$$ab \in \left\{ \mathfrak{M}(b)\frac{\sigma_{YX}\sigma_{YM} \pm \sqrt{\sigma_Y^2 \sigma_X^2 - \sigma_{YX}^2}\sqrt{\sigma_M^2 \sigma_Y^2 - \sigma_{YM}^2}}{\sigma_X^2 \sigma_Y^2} \right\}. \tag{A20}$$

The maximum possible indirect effect is obtained by the product of $\mathfrak{M}(a)$ and $\mathfrak{M}(b)$:

$$\mathfrak{M}(ab) = \mathfrak{M}(a)\mathfrak{M}(b). \tag{A21}$$

Received September 20, 2009
Revision received August 20, 2010
Accepted August 30, 2010 ■