Psychological Methods

1082-989X/17/$12.00 http://dx.doi.org/10.1037/met0000165

A Novel Measure of Effect Size for Mediation Analysis

Mark J. Lachowicz and Kristopher J. Preacher
Vanderbilt University

Ken Kelley
University of Notre Dame

Abstract

Mediation analysis has become one of the most popular statistical methods in the social sciences. However, many currently available effect size measures for mediation have limitations that restrict their use to specific mediation models. In this article, we develop a measure of effect size that addresses these limitations. We show how modification of a currently existing effect size measure results in a novel effect size measure with many desirable properties. We also derive an expression for the bias of the sample estimator for the proposed effect size measure and propose an adjusted version of the estimator. We present a Monte Carlo simulation study conducted to examine the finite sampling properties of the adjusted and unadjusted estimators, which shows that the adjusted estimator is effective at recovering the true value it estimates. Finally, we demonstrate the use of the effect size measure with an empirical example. We provide freely available software so that researchers can immediately implement the methods we discuss. Our developments here extend the existing literature on effect sizes and mediation by developing a potentially useful method of communicating the magnitude of mediation.

Translational Abstract

An effect size is often used to translate a result obtained from a specific study into a metric that is independent of arbitrary characteristics of the study design (e.g., variable scales), making it easier for researchers to communicate the importance of their results and compare them with those obtained from other studies. The purpose of this research is to propose such an effect size for mediation analysis. Mediation analysis is used to examine the processes through which a predictor has an effect on an outcome through intervening variables called mediators. The component of an effect transmitted via a mediator is known as an indirect effect. Although indirect effects are commonly reported, effect size measures for them have yet to be firmly established. We show that our proposed measure is an attractive option for several reasons, but most importantly that it (a) has an intuitive interpretation, (b) quantifies the indirect effect independent of arbitrary design choices, and (c) can be used to draw valid inferences for sample sizes and effect magnitudes common in applied research. We then demonstrate the application and interpretation of the effect size using real data, and provide freely available software so researchers may immediately use the measure in their research.

Keywords: mediation analysis, effect size, indirect effect

Scholars in many fields have long advocated that researchers move away from null hypothesis significance tests (NHSTs) and *p* values as the primary source of support for their hypotheses (Cohen, 1994; Greenland & Poole, 2013; Keuzenkamp & Magnus, 1995; Morrison & Henkel, 1970; Wilkinson & American Psychological Association Task Force on Statistical Inference, 1999). A primary criticism of NHSTs is that *p* values provide no information about the size or importance of effects, only the likelihood of obtaining an effect as large or larger than that actually obtained under a given null hypothesis (with appropriate assumptions met). The push for effect sizes is mandated in the most recent American Psychological Association (APA) publication manual (APA, 2010), in which it is stated that NHSTs are "a starting point and that additional reporting elements such as effect sizes, confidence intervals, and extensive description are needed to convey the most complete meaning of the results" (p. 33).

Effect size measures have been developed and routinely employed for quantifying the magnitude of effects for many classic statistical methods (Cohen, 1988; Kirk, 1996). These effect size measures include Cohen's *d*, correlations, regression coefficients, and R^2, among many others. However, there are some statistical methods for which consensus has not been reached regarding the most appropriate effect size measures. One of the most notable among these methods without consensus, and one of the most important methods in modern research, is mediation analysis.

Mediation analysis is the study of the potential pathways through which a predictor (independent) variable has an effect on an outcome (dependent) variable. These pathways can transmit the effect to the dependent variable at least partially through interven-

Mark J. Lachowicz and Kristopher J. Preacher, Department of Psychology & Human Development, Vanderbilt University; Ken Kelley, Information Technology, Analytics, and Operations, University of Notre Dame.

We thank Sonya Sterba and James H. Steiger, whose insightful comments and suggestions substantially improved the article.

Correspondence concerning this article should be addressed to Mark J. Lachowicz, Quantitative Methods Program, Department of Psychology & Human Development, Vanderbilt University, Peabody #552, 230 Appleton Place, Nashville, TN 37203. E-mail: mark.j.lachowicz@vanderbilt.edu

ing variables called *mediators*. Because of the potential for this framework to aid the understanding of processes, mediation analysis has become one of the most popular statistical methods used in applied research. Whereas much progress in mediation analysis has been made regarding statistical inference for mediation effects, less progress has been made in developing effect size measures for mediation.

Our purpose in this article is to propose a novel effect size measure for mediation analysis to address existing deficiencies in the current state of the art in mediation analysis. The article has six main parts. In the first section, we review mediation analysis. In the second section, we review effect size, particularly effect size measures developed specifically for mediation analysis. In the third section, we discuss properties of the sample estimator of the effect size and propose an estimator that adjusts for sample bias. The fourth section details the methods and results of a Monte Carlo simulation conducted to examine the finite sampling properties of the adjusted and unadjusted estimators. The fifth section contains an empirical example that demonstrates the use of the estimators. The sixth section is a summary of findings, limitations, and future directions.

Mediation Analysis

Figure 1 provides an example of a path diagram for a classic single-level, three-variable mediation model (e.g., Baron & Kenny, 1986; MacKinnon, 2008). The *total effect* of a predictor X on an outcome Y in the population is given by the regression equation

$$Y = d_{YX} + B_{YX}X + e_{Y \cdot X}, \tag{1}$$

where d_{YX} is the intercept, B_{YX} is the linear slope relating X to Y, and $e_{Y \cdot X}$ is the residual error term, where $e_{Y \cdot X} \sim N(0, \sigma^2_{e_{Y \cdot X}})$ (i subscripts for random variables are omitted for convenience). How the presumed effect of X is transmitted to Y via intervening variables (i.e., mediators) is examined using a system of two additional linear regressions. The relationship between X and the mediator M in the population is expressed as

$$M = d_{MX} + B_{MX}X + e_{M \cdot X}, \tag{2}$$

where d_{MX} is the intercept, B_{MX} is the linear slope relating X to M, and $e_{M \cdot X}$ is the error term, where $e_{M \cdot X} \sim N(0, \sigma^2_{e_{M \cdot X}})$. The equation relating Y to both X and M in this population mediation model is expressed as

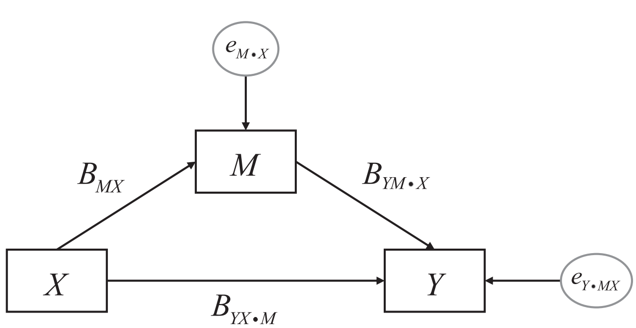

Figure 1. Path diagram for a three-variable mediation model.

$$Y = d_{Y \cdot MX} + B_{YX \cdot M}X + B_{YM \cdot X}M + e_{Y \cdot MX}, \tag{3}$$

where $d_{Y \cdot MX}$ is the intercept, $B_{YM \cdot X}$ and $B_{YX \cdot M}$ are the slope coefficients of Y regressed on M controlling for X and X controlling for M, respectively, and $e_{Y \cdot MX}$ is the error term, where $e_{Y \cdot MX} \sim N(0, \sigma^2_{e_{Y \cdot MX}})$. It is assumed that the errors $e_{M \cdot X}$ and $e_{YM \cdot X}$ are uncorrelated, meaning there is no unmodeled variable influencing the relationship between M and Y (i.e., no omitted confounders or model misspecification). The effect of X on Y independent of M, denoted $B_{YX \cdot M}$, is referred to as the *direct effect*. The effect of X on Y transmitted via M is referred to as the *indirect effect*, which is calculated as the product of B_{MX} from Equation 2 and $B_{YM \cdot X}$ from Equation 3. These population parameters are estimated by their respective unbiased sample estimators $\hat{B}_{MX}$, $\hat{B}_{YX.M}$, and $\hat{B}_{YM.X}$. Much of the progress in the methods literature for mediation analysis has been concerned with inference regarding the indirect effect, such that several methods are currently available for obtaining confidence intervals (CIs; e.g., bootstrap CIs [Bollen & Stine, 1990; MacKinnon, Lockwood, & Williams, 2004; Shrout & Bolger, 2002], Monte Carlo CIs [MacKinnon et al., 2004], Bayesian credible intervals [Yuan & MacKinnon, 2009], and a method of constructing CIs based on the distribution of product terms [MacKinnon, Fritz, Williams, & Lockwood, 2007; Tofighi & MacKinnon, 2011]).

It is important to note that although mediation analysis provides a means for researchers to investigate causal processes, the model outlined in Equations 1 to 3 cannot prove causality. For example, Cohen, Cohen, West, and Aiken (2003) outline the fundamental conditions necessary for making causal inferences including association, temporal precedence of cause before effect, and isolation of the causal effect from confounding variables. The issue of confounding of the causal effect in mediation analysis has received substantial attention in the methodological literature over the last decade (e.g., Imai, Keele, & Tingley, 2010; Pearl, 2014; Vanderweele, 2015). This research has advanced our current understanding and application of mediation models, with such noteworthy contributions as the identification of the necessary assumptions for the indirect effect to be considered causal, the derivation of the biases in causal effect estimates if assumptions are violated, and the development of methods to evaluate estimates for robustness to these violations. In terms of the mediation model in Figure 1, the additional assumptions required for the indirect effect to be considered a causal effect are (a) constant effect, meaning that the effect of M on Y does not vary across levels of X (i.e., no interaction of X and M); and (b) sequential ignorability, meaning the effects of X on M and X on Y, as well as the effect of M on Y, are not confounded by omitted variables or model misspecification. The constant effect assumption may be relaxed by modeling the interaction of X and M, resulting in indirect effects conditional on levels of X. Although the sequential ignorability assumption cannot be so easily relaxed, sensitivity analysis, in which changes in the magnitude and precision of the proposed causal effect are examined assuming various associations with hypothetical confounders, can provide important evidence to support the claim of a causal effect.

Effect Size

Broadly speaking, an *effect size* is defined as the quantification of some phenomenon of interest to address a specific research question (K. Kelley & Preacher, 2012). Although any sample statistic can function as a measure of effect size in the proper context, there are several properties a statistic should have in order for it to be considered a useful measure of effect size (K. Kelley & Preacher, 2012; Preacher & Kelley, 2011). First, the statistic should have an interpretable scale. Whereas some measures of effect size are useful in their original metrics (e.g., mean differences on an established measure; Baguley, 2009), it is often useful to standardize effect size measures. Standardization makes results more comparable across studies by removing the metric of one or more variables. Second, it should be possible to construct CIs based on the statistic's sampling distribution. Third, the statistic should be unbiased, consistent, and efficient. Fourth, the population effect size should be independent of sample size. In addition to these criteria, Wen and Fan (2015) assert that, with all else held constant, an effect size measure should be a monotonic function in raw or absolute value of the effect it quantifies. This collection of properties ensures the effect size measure is a good estimator of the population value it estimates, allows for the comparison of results across studies, facilitates meta-analyses, and facilitates sample size calculations for future studies (Wilkinson & American Psychological Association Task Force on Statistical Inference, 1999).

Standardized mean differences, correlation coefficients, and proportion of variance measures are appropriate effect size measures for many questions of interest in traditional research in psychology and other disciplines (Cohen, 1988). However, there are some nonstandard regression-based methods, notably mediation analysis, for which these measures are not sufficient. In mediation analysis, the primary statistic of interest is generally the indirect effect, which, as we discussed, is a product of regression coefficients. Whereas the individual regression coefficients themselves are established measures of effect size, the effect size for the indirect effect is not captured by those measures. For example, consider an indirect effect for a three-variable mediation model in which the effect size of one coefficient is considered "large" by some disciplinary standard and the effect size of the other is considered "small." It is unclear, then, how the effect size of the indirect effect itself should be quantified—effect size measures of the component coefficients B_{MX} and $B_{YM \cdot X}$ were designed to quantify linear relationships defined by a single equation. The most basic representation of mediation, however, involves at least three variables and two equations. Correspondingly, effect size measures for simpler models do not adequately capture the more complex effect in mediation analysis.

Several effect size measures have been proposed to quantify the size of the indirect effect in the context of mediation analysis, in which these measures take qualitatively different approaches. The first effect size measures proposed for indirect effects were ratios, which compare the indirect effect with the total effect (P_M; Alwin & Hauser, 1975) or compare the indirect effect with the direct effect (R_M; Sobel, 1982). Whereas these measures are easy to implement and have some intuitive appeal, simulation studies have identified serious problems and their use is generally not recommended (MacKinnon, Warsi, & Dwyer, 1995; Preacher & Kelley, 2011). Despite these limitations, ratio measures have been the most commonly reported effect sizes for indirect effects, in large part due to a lack of viable alternatives. However, there have been several more recent efforts to address this gap in the literature.

Several methodologists have proposed standardized indirect effects as effect size measures for mediation analysis (Alwin & Hauser, 1975; Cheung, 2009; MacKinnon, 2008; Preacher & Hayes, 2008). For a three-variable mediation model, the indirect effect can be standardized in two ways: (a) standardization by the scale of only Y, and (b) standardization by the scales of Y and X. Standardization by the scale of only Y is sensible if the metric of X is meaningful but the metric of Y is not. For example, if Y is continuous but X is an indicator of group membership, partial standardization would be appropriate because the metric of X (a unit change in X is a change in group membership) is already meaningful and comparable across studies. However, partial standardization would make across-study comparisons difficult if X were continuous and the scale of X differed across populations. For these models, complete standardization of the indirect effect by the scales of both X and Y is sensible. Referred to as the "index of mediation" by Preacher and Hayes (2008), the completely standardized indirect effect is invariant to linear transformations of X, M, or Y, making effects comparable across studies in which the scales X and Y differ. The completely standardized indirect effect also has an intuitive interpretation as the expected standard deviation change in Y for a one standard deviation increase in X through M. Finally, the standardized indirect effect has the good statistical properties of unbiasedness and consistency (Cheung, 2009). For these reasons, the partially or completely standardized indirect effect is an attractive effect size measure for many mediation models.

More recently, Kraemer (2014) proposed an effect size measure for indirect effects in the context of randomized clinical trials (RCTs). Referred to as *MedES*, this measure quantifies the indirect effect as the difference between the overall success rate difference between treatment and control conditions, and the success rate difference between treatment and control had the relationship between the mediator and outcome been severed. Although a potentially desirable effect size measure in the context of a RCT, the requirement that the independent variable be binary limits the applicability of *MedES* because many studies consider continuous predictors.

In their review of effect size in mediation analysis, Preacher and Kelley (2011) proposed two new measures: (a) κ^2, and (b) Γ. The effect size κ^2 was defined as the ratio of the observed indirect effect to the maximum possible indirect effect that *could have been* observed given the study design conditional on observed statistics. Wen and Fan (2015) identified two properties of κ^2 that suggest it has limitations as a general effect size measure. The first such limitation is that κ^2 is not a strictly monotonic function in raw or absolute value of the indirect effect. This means that with all else held equal (i.e., sample size, total variances, residual variances, total effect), an increase in the indirect effect does not necessarily correspond to an increase in the effect size. Second, although the maximum $B_{YM \cdot X}$ can be found for a given set of data, the maximum possible indirect effect is theoretically infinite. Although the underlying concept of κ^2 is appealing for an effect size measure, deficiencies limit its theoretical usefulness. The effect size Γ was defined as the ratio of the variance explained in M by X and the variance explained in Y jointly by M and X to the total variability of M and Y. This was an extension of an effect size measure (M_{BM}) proposed by Berry and Mielke (2002) for use in multivariate multiple regression. Some desirable characteristics of Γ are that it has a meaningful interpretation as a proportion bounded by 0 and 1, it is independent of N, and bootstrap CIs can be estimated. However, Γ can return a nonzero effect size when the null hypothesis is true (i.e., $B_{MX}B_{YM \cdot X} = 0$). This is possible because Γ was derived under the assumption that the null hypothesis is true when X explains no variance in either M or Y, and M explains no variance in Y. However, an indirect effect of zero can also occur when X explains variance in M, but M and X explain no variance in Y (i.e., $B_{YM \cdot X} = 0$). This suggests that the null hypothesis assumed for Γ is not the only null hypothesis under which the indirect effect is zero. Formulations for these additional null hypotheses may need to be incorporated to account for all parameter combinations that correspond to an indirect effect of zero.

R^2 measures in mediation analysis are intended to quantify the proportion of variance explained in the outcome that can be attributed to both the predictor and the mediator but to neither alone (de Heus, 2012; Fairchild, MacKinnon, Taborga, & Taylor, 2009; MacKinnon, 2008). The reason this variance explained jointly by M and X is a R^2 measure for the indirect effect can be understood by considering the three sources of variance in Y explained by M and X: (a) variance attributable to M independent of X, (b) variance attributable to X independent of M, and (c) variance attributable to X and M jointly. Because a mediation model represents a decomposition of the total effect of X on Y, any variance in Y not attributable to X either uniquely or jointly is irrelevant, which would exclude the first source of variance in Y listed above as relevant to an R^2 measure for the indirect effect. Of the remaining two sources of variance, the second source is consistent with the definition of the direct effect, or the effect of X on Y controlling for M, therefore leaving the only source of variance in Y relevant to a R^2 measure for the indirect effect as the variance attributable to X and M jointly. One such approach for quantifying this variance was developed by Fairchild and colleagues (2009), in which the variance in Y jointly explained by M and X is

$$R^2_{med} = \rho^2_{YM} - (R^2_{Y \cdot MX} - \rho^2_{YX}), \tag{4}$$

where ρ^2_{YM} is the squared correlation between M and Y, ρ^2_{YX} is the squared correlation between X and Y, and $R^2_{Y \cdot MX}$ is the squared multiple correlation of Y with M and X. How R^2_{med} isolates the variance explained in Y jointly by M and X can be understood by considering the components of variance in Y explained by M (i.e., variance components of ρ^2_{YM}). The quantity ρ^2_{YM} consists of two variance components from the sources previously described: (a) the desired joint variance explained, and (b) variance in Y explained by M independent of X. The former joint variance component is the difference between the total variance explained, ρ^2_{YM}, and the latter variance component, which is equivalent to the difference between the total variance in Y explained by M and X ($R^2_{Y \cdot MX}$) and the total variance in Y explained by X alone (ρ^2_{YX}).[1] A significant limitation of this measure is that, like Preacher and Kelley's (2011) residual-based index (i.e., Γ), it returns nonzero effect sizes when the indirect effect is zero. In addition, as demonstrated by Preacher and Kelley, R^2_{med} is neither a monotonically increasing function of the indirect effect in either raw or absolute value, nor is it bounded by 0 and 1, properties that make interpretation exceedingly difficult. Two additional R^2 effect size measures were proposed by MacKinnon (2008) for the simple three-variable mediation model in Equations 1 to 3:

$$R^2_{4.6} = \rho^2_{MX} \times \rho^2_{YM \cdot X} \tag{5}$$

and

$$R^2_{4.7} = \frac{\rho^2_{MX} \times \rho^2_{YM \cdot X}}{R^2_{Y \cdot MX}}, \tag{6}$$

where ρ^2_{MX} is the squared correlation of M and X, and $\rho^2_{YM \cdot X}$ is the squared partial correlation of Y and M adjusting for X. Equation 6 represents a scaling of Equation 5 by the reciprocal of the total proportion of variance in Y accounted for by M and X. More recently, de Heus (2012) proposed a modification of the measure in Equation 5,

$$R^2_{DH} = \rho^2_{MX} \times \rho^2_{Y(M \cdot X)}, \tag{7}$$

where $\rho^2_{Y(M \cdot X)}$ is the squared semipartial correlation of Y and M adjusting for X. Equation 7 represents a scaling of the partial correlation coefficient by the proportion of variance in Y accounted for by X. Because ρ^2_{MX}, $\rho^2_{YM \cdot X}$, and $\rho^2_{Y(M \cdot X)}$ are bounded by 0 and 1, the R^2 formulations in Equations 5 to 7 are also bounded by 0 and 1. However, like R^2_{med}, the R^2 formulations in Equations 5 to 7 are not monotonically increasing functions of the indirect effect in raw or absolute value. This is shown in Panels A to C of Figure 2, for which population effect sizes from Equations 5 to 7 were calculated, respectively, for 5,000 randomly generated positive definite correlation matrices using the uniform correlation matrix method proposed by Botha, Shapiro, and Steiger (1988) (results in Panel D will be discussed in a later section). Specifically, it can be seen in Panels A to C that a specific value of an indirect effect has multiple corresponding values of effect size. Most commonly used effect size measures (e.g., Cohen's d, Cohen's f, R^2) are monotonically increasing functions of the quantity of interest in raw or absolute value at the population level. The conceptions of effect size in Equations 5 to 7 are not monotonically increasing functions of the indirect effect because the indirect effect is theoretically unbounded, and the method by which Equations 5 to 7 constrain the effect size to be bounded by 0 and 1 requires standardized indirect effects greater than 1 to be scaled ≤ 1.

In summary, many existing effect size measures for mediation analysis have properties that limit their use. This suggests that to address the limitations of current effect size measures, an effect size measure for mediation analysis, in addition to having the properties outlined in Preacher and Kelley (2011) and Wen and Fan (2015), should return an effect size of zero when the indirect effect is zero. There is a clear need for an effect size measure for mediation analysis that satisfies these properties, as, at present, for one of the most popular models in modern social science and the call for effect sizes to be used to illustrate the size of effects, an ideal effect size for mediation models is needed. Our primary aim was to develop such an effect size measure so that mediation analysis can follow the recommendation of the APA and others to report effect sizes in addition to hypothesis tests.

[1] Alternatively, how R^2_{med} isolates the variance in Y explained jointly by M and X can be understood by considering first ρ^2_{YX} and ρ^2_{YM}. Both consist of variance in Y explained jointly by M and X, and distributing the negative sign in Equation 4 shows that these quantities are summed, meaning the joint variance explained is counted twice. Therefore, subtracting $R^2_{Y \cdot MX}$ from this sum yields the area of joint overlap.

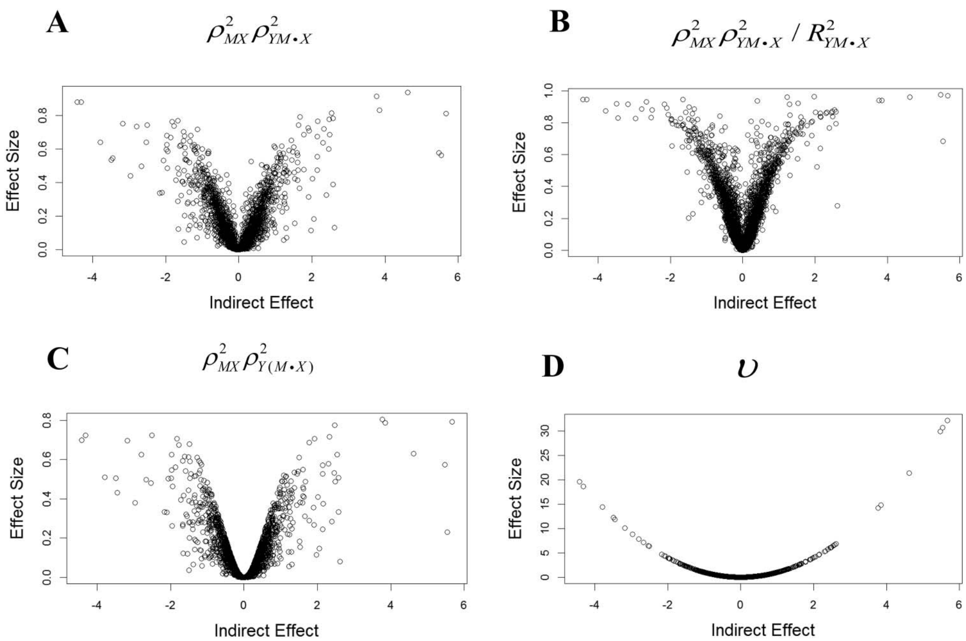

Figure 2. Plots of R^2 effect sizes measures versus the indirect effect for a three-variable mediation model for 5,000 indirect effects. Effect sizes in Panels A and B refer to Equations 5 and 6 (MacKinnon, 2008); effect size in Panel C refers to Equation 7 (de Heus, 2012); and effect size in Panel D refers to Equation 14 ($\tilde{\upsilon}$).

A Novel Effect Size Measure for Mediation Analysis

The effect size measure we propose is a modification of the R^2_{med} measure from Equation 4. The R^2_{med} approach for obtaining shared variance is based on a multiple regression framework in which a single equation represents the relationships among variables. For the mediation model presented in Figure 1, the proportion of variance accounted for in Y by X and M jointly according to Fairchild et al. (2009) is a function of the squared zero-order correlations ρ^2_{YM} and ρ^2_{YX}, and the squared multiple correlation $R^2_{Y \cdot MX}$. Although this method is appropriate for quantifying variance explained jointly in the multiple regression context, the model in Figure 1 makes additional assumptions about the relationships among variables that cannot be modeled in a single multiple regression equation. The modification of R^2_{med} that we propose results in an explained variance effect size measure consistent with the assumptions of a path analysis framework.

As shown in Equations 1 to 3, a mediation model requires a system of equations specifying the causal ordering of variables (i.e., X causes M, X and M cause Y). This causal ordering of variables imposes the assumptions that Y is dependent on M, and that M and Y are mutually dependent on X, given that several additional assumptions hold (temporal precedence of variables, isolation of effects, association, etc.). As in the structural equation modeling literature, X here is referred to as an *exogenous* variable, and M and Y are *endogenous*. This assumption implies that the zero-order correlation between M and Y has two components: (a) the correlation between M and Y independent of X, and (b) the correlation due to the mutual dependence of M and Y on X (Duncan, 1970; Wright, 1960). The conditional correlation is often referred to as *true correlation*, and the correlation due to mutual dependence is referred to as *spurious correlation* (Blalock, 1962; Dillon & Goldstein, 1984; Simon, 1954). For X to have an indirect effect on Y through M, part of the correlation between Y and M must be true correlation, otherwise M does not cause a change in Y when controlling for X. It is important to note that this assumes a correctly specified model with the appropriate assumptions being met, so the spurious correlation described here is distinct from spurious correlation induced by the omission of confounding variables.

To illustrate how spurious correlation is quantified in mediation analysis, consider again the model in Figure 1. If X has nonzero effects on both M and Y, the zero-order correlation between Y and M is due partly to these associations with X. Specifically, the correlation between Y and M (i.e., via path tracing rules) is decomposed as

$$\rho_{YM} = \beta_{YM \cdot X} + \beta_{YX \cdot M}\beta_{MX}, \tag{8}$$

where $\beta_{YM \cdot X}$, $\beta_{YX \cdot M}$, and β_{MX} are standardized regression coefficients. True correlation is quantified by $\beta_{YM \cdot X}$, and spurious correlation is quantified by $\beta_{YX \cdot M}\beta_{MX}$. In circumstances in which there is no indirect effect, it is either the case that the true correlation component $\beta_{YM \cdot X}$ is zero, and $\rho_{YM} = \beta_{YX \cdot M}\beta_{MX}$, or the spurious correlation component $\beta_{YX \cdot M}\beta_{MX}$ is zero (i.e., $\beta_{MX} = 0$), and $\rho_{YM} = \beta_{YM \cdot X}$. Alternatively, when there is no direct effect, the spurious correlation component $\beta_{YX \cdot M}\beta_{MX}$ is zero (i.e., $\beta_{YX \cdot M} = 0$, and $\rho_{YM} = \beta_{YM \cdot X}$). In other words, ρ_{YM} cannot be used to distinguish between scenarios in which indirect effects are present or absent.

The implications of distinguishing true and spurious correlation when quantifying the variance explained by the indirect effect can be seen by considering the behavior of R^2_{med} when indirect effects are absent (i.e., $\beta_{MX} = 0$, $\beta_{YM \cdot X} = 0$, or $\beta_{MX} = \beta_{YM \cdot X} = 0$).

Inserting the decomposition of ρ_{MY} in Equation 8 into Equation 4, and using the identities $\rho_{YX} = \beta_{YX}$, $\rho_{MX} = \beta_{MX}$, and

$$R^2_{Y \cdot MX} = \beta^2_{YX} + \beta^2_{YM \cdot X}(1 - \beta^2_{MX}), \tag{9}$$

the variance in Y explained jointly by M and X when $\beta_{MX} = 0$ is

$$\begin{aligned} R^2_{med} &= (\beta_{YM \cdot X} + \beta_{YX \cdot M}\beta_{MX})^2 - \left[\beta^2_{YX} + \beta^2_{YM \cdot X}(1 - \beta^2_{MX}) - \beta^2_{YX}\right] \\ &= \beta^2_{YM \cdot X} - \beta^2_{YM \cdot X} + \beta^2_{YX} - \beta^2_{YX} = 0 \end{aligned} \tag{10}$$

As previously noted, when $\beta_{MX} = 0$, ρ_{MY} consists of only true correlation, and the resulting effect size of zero is consistent with an indirect effect of zero. R^2_{med} is also consistent with an indirect effect of zero when $\beta_{MX} = \beta_{YM \cdot X} = 0$

$$\begin{aligned} R^2_{med} &= (\beta_{YM \cdot X} + \beta_{YX \cdot M}\beta_{MX})^2 - \left[\beta^2_{YX} + \beta^2_{YM \cdot X}(1 - \beta^2_{MX}) - \beta^2_{YX}\right] \\ &= \beta^2_{YX} - \beta^2_{YX} = 0 \end{aligned} \tag{11}$$

Here, there is no true or spurious correlation, and, thus, the zero-order correlation is also zero. However, the variance in Y explained by X via the indirect effect when $\beta_{YM \cdot X} = 0$ is

$$\begin{aligned} R^2_{med} &= (\beta_{YM \cdot X} + \beta_{YX \cdot M}\beta_{MX})^2 - \left[\beta^2_{YX} + \beta^2_{YM \cdot X}(1 - \beta^2_{MX}) - \beta^2_{YX}\right] \\ &= \beta^2_{YX \cdot M}\beta^2_{MX} + \beta^2_{YX} - \beta^2_{YX} \\ &= \beta^2_{YX \cdot M}\beta^2_{MX}. \end{aligned} \tag{12}$$

The resulting nonzero effect size in this scenario is inconsistent with an indirect effect of zero. As previously noted, when $\beta_{YM \cdot X} = 0$, ρ_{MY} consists of only spurious correlation, meaning the joint explained variance quantified by R^2_{med} in this scenario is not attributable to the indirect effect.

Because variance explained by spurious correlation is not associated with the indirect effect, it follows that a variance explained effect size measure for the indirect effect should not incorporate this correlation. The variance in Y explained by M can be adjusted for spurious correlation by subtracting $\beta_{YX \cdot M}\beta_{MX}$ from the zero-order correlation, which is equivalent to replacing the squared correlation between Y and M with the squared standardized regression coefficient $\beta_{YM \cdot X}$:

$$(\rho_{YM} - \beta_{MX}\beta_{YX \cdot M})^2 = \beta^2_{YM \cdot X}. \tag{13}$$

Inserting the result from Equation 13 into the R^2_{med} formula in Equation 4 gives an adjusted version of R^2_{med}, which we will refer to hereafter as upsilon (υ), denoted at the population level as

$$\upsilon = (\rho_{YM} - \beta_{MX}\beta_{YX \cdot M})^2 - (R^2_{Y \cdot MX} - \rho^2_{YX}). \tag{14}$$

The effect size υ is a novel method for quantifying explained variance in mediation analysis, and is interpreted as *the variance in Y accounted for jointly by M and X that corrects for spurious correlation induced by the ordering of variables*.

A circumstance in a three-variable mediation model in which R^2_{med} would not need to be adjusted is when the spurious correlation is zero (i.e., $\beta_{MX}\beta_{YX \cdot M} = 0$). This circumstance is possible if $\beta_{MX} = 0$, $\beta_{YX \cdot M} = 0$, or $\beta_{MX} = \beta_{YX \cdot M} = 0$. Because there is no indirect effect when $\beta_{MX} = 0$, the only circumstance under which spurious correlation would not need to be accounted for is when the direct effect is zero (i.e., $\beta_{YX \cdot M} = 0$). That is, R^2_{med} and υ are equivalent when the effect of X on Y is completely accounted for by M. Therefore, it is not necessary to employ R^2_{med} because, under the only circumstances in which the measure is appropriate, R^2_{med} is a special case of υ.

Adjusting R^2_{med} for spurious correlation (i.e., υ) results in a noteworthy interpretation for the variance explained in Y jointly by M and X. Consider υ from the previous three-variable mediation example in Equation 14, which can be reexpressed as

$$\upsilon = \beta^2_{YM \cdot X} - (R^2_{Y \cdot MX} - \rho^2_{YX}). \tag{15}$$

Equation 15 can be reexpressed in terms of standardized regression coefficients using the identities for $R^2_{Y \cdot MX}$ in Equation 9 and $\rho^2_{YX} = \beta^2_{YX}$

$$\begin{aligned} \upsilon &= \beta^2_{YM \cdot X} - \left[\beta^2_{YX} + \beta^2_{YM \cdot X}(1 - \beta^2_{MX}) - \beta^2_{YX}\right] \\ &= \beta^2_{YM \cdot X} - \beta^2_{YM \cdot X}(1 - \beta^2_{MX}) \\ &= \beta^2_{MX}\beta^2_{YM \cdot X}. \end{aligned} \tag{16}$$

In other words, for three-variable mediation, *the variance in Y accounted for jointly by M and X adjusting for the ordering of variables is equivalent to the squared standardized indirect effect*. This result suggests that to obtain an R^2 effect size measure for the indirect effect in three-variable mediation, one estimates the unstandardized indirect effect, standardizes the estimated indirect effect by the variance of X and variance of Y, and squares the completely standardized indirect effect.

de Heus (2012) was the first to introduce the concept of the squared standardized indirect effect as a measure of explained variance, and explored the consequences of modeling the direct effect on the interpretation of the variance explained by the indirect effect. de Heus posited that, because the squared standardized total effect quantified the variance explained in Y by X, measures of variance explained by the direct and indirect effects could be obtained by substituting the squared standardized total effect with the squared sum of the standardized indirect and direct effects. This results in three components of explained variance: (a) unique variance explained by the indirect effect (β^2_{IND}), (b) unique variance explained by the direct effect (β^2_{DIR}), and (c) a component representing variance explained not attributable to either the direct or indirect effect alone. The results of this study confirm that β^2_{IND} indeed represents the unique variance explained by the indirect effect. de Heus proposed two methods for quantifying the variance explained by the direct and indirect effects that either ignores the joint component of variance explained by the effects, or assigns the joint overlap to either the direct or indirect effect depending on which is of primary interest to the researcher. Because an effect size measure for the indirect effect is of primary interest in the present research (i.e., in quantifying the size of the indirect effect in a mediation model), the components of variance explained by the indirect effect that result from the two methods are conceptually and mathematically equivalent. In other words, the method of assigning the joint variance explained by the direct and indirect effects is inconsequential for an effect size measure designed specifically for the indirect effect.

A factor to consider when interpreting υ is the behavior of standardized multiple regression coefficients. Although it may appear that υ is interpretable as a proportion of explained variance, it is important to realize that standardized regression coefficients can be greater than 1 (Cohen et al., 2003; Jöreskog, 1999), so the squared standardized regression coefficients should not be considered proportions. That is, squared path coefficients do not function as proportions as do squared correlation coefficients (or squared partial or semipartial correlation coefficients), and, therefore, although υ is a measure of variance explained, it is not interpretable as a proportion.

Of the effect size measures for mediation analysis presented in the previous section, υ most closely resembles the R^2 effect size measures in Equations 5 to 7 (i.e., $\rho^2_{MX} \times \rho^2_{YM \cdot X}$, $\rho^2_{MX} \times \rho^2_{YM \cdot X}/R^2_{Y \cdot MX}$, and $\rho^2_{MX} \times \rho^2_{Y(M \cdot X)}$). Although the R^2 effect size measures in Equations 5 to 7 are bounded by 0 and 1 and can be interpreted as proportions, υ has a key advantage as an R^2 effect size measure. This advantage becomes apparent when suppression is evident. In mediation analysis, suppression occurs when the addition of a mediator strengthens the relationship between the predictor and outcome (MacKinnon, Krull, & Lockwood, 2000; Rucker, Preacher, Tormala, & Petty, 2011), such that the direct and indirect effects can have magnitudes greater than the total effect when suppression is evident. As previously noted, the R^2 effect size measures in Equations 5 to 7 are always bounded by 0 and 1. When suppression is not evident, υ is also bounded by 0 and 1.[2] In addition, all R^2 effect size measures are monotonic functions in absolute value of the indirect effect when there is no suppression, differing only in scale (i.e., there is a one-to-one relationship between the effect size and the standardized indirect effect in the population). However, υ remains a monotonic function of the indirect effect in the population when suppression is evident (see Panel D of Figure 2). This is a desirable property for an effect size measure in mediation analysis because a larger effect size directly translates to a larger indirect effect, whereas for a nonmonotonic effect size it would be unclear if a larger effect size were due to a larger indirect effect, suppression effects, or some combination thereof.

Properties of υ

The measure υ has several desirable properties for an effect size for quantifying the indirect effect. First, υ is standardized, so it is scale invariant to linear transformations of X, M, and/or Y. Second, υ is a type of variance explained measure, which is a family of statistics that are widely used in applied research settings and have intuitive interpretations. Third, υ is not dependent on sample size. Fourth, υ is a monotonic function in absolute value of the standardized indirect effect. Finally, υ may be used for models with continuous or binary X because the standardized coefficient of a binary variable in simple linear regression is equivalent to a Pearson correlation coefficient, and the squared correlation coefficient is the variance explained in the outcome by the binary variable. Taken together, these important properties make υ a promising effect size measure for mediation analysis with a highly desirable set of properties.

In order to appropriately interpret the magnitude of an effect, it is often useful for researchers to consult benchmarks for effect size measures. Although not universally applicable, effect size benchmarks often facilitate the understanding and communication of results by allowing researchers to compare their estimates with established references for small, medium, and large effects. Although a continuing subject of debate in the social science literature, the most commonly cited of these benchmarks are those proposed by Cohen (1988) for standardized regression coefficients (small = 0.14, medium = 0.39, and large = 0.59) and proportions of explained variance (small = 2%, medium = 15%, and large = 25%). Because υ is being developed here, there is not yet a body of literature from which to draw appropriate benchmarks.

Nevertheless, because υ is a measure of explained variance, there is justification for evaluating υ in terms of Cohen's benchmarks for proportions of variance explained. Consider the example of a three-variable mediation population model in which the direct effect is zero. The key property of this model to note is that the total and indirect effects are equivalent (i.e., $\beta_{MX}\beta_{YM \cdot X} = \beta_{YX}$). When effects are standardized as we have assumed here, the total effect is a standardized regression coefficient, the magnitude of which can be appropriately compared with Cohen's benchmarks. An equivalent comparison is to square the standardized total effect, which yields the proportion of variance in Y explained by X. This R^2 can also be compared with Cohen's benchmarks. Because of the equivalence of the total and indirect effects when the direct effect is zero, υ for this model would result in the same amount of variance explained in Y by X through M as would be explained in Y directly by X without considering M. The equivalence of υ and total R^2 in this model suggests the effect sizes should have the same interpretation in terms of variance explained in Y, and Cohen's benchmarks for small, medium, and large proportions of explained variance are applicable to υ. However, we also stress that Cohen's benchmarks are conventions and not universally applicable in all research settings, so researchers should carefully consider prior research and the context of their study when interpreting the magnitudes of effects.

Sample Estimator of υ

The effect size properties described thus far apply to υ in the population. The remaining properties outlined in Preacher and Kelley (2011) apply to the sample estimator of υ (i.e., bias, consistency, efficiency, CIs). Because the population υ is equivalent to the squared standardized indirect effect for three variable mediation models, a natural choice for the estimator is the sample squared standardized indirect effect $\hat{\upsilon} = \hat{\beta}^2_{MX}\hat{\beta}^2_{YM \cdot X}$.

It is not always true that the expected value of a sample estimator obtained by inserting sample values for the population analog is equivalent to its population parameters. Perhaps the most notable example of this fact is sample estimators of variance parameters. It has long been acknowledged in the methodological literature that R^2 estimates are positively biased (Ezekiel, 1930; T. L. Kelley, 1935; Olkin & Pratt, 1958). Several formulas have been proposed to adjust R^2 for this bias in ANOVA and multiple regression (see Yin & Fan, 2001, for a comprehensive review).

Because υ is an explained variance measure, it is expected that $\hat{\upsilon}$ is positively biased, meaning that $\hat{\upsilon}$ tends to be larger than the corresponding population value. If $\hat{\upsilon}$ were unbiased, it could be shown that $E[\hat{\upsilon}] = \upsilon$, or, alternatively, that the expected value of the sample squared standardized indirect effect is the population squared standardized indirect effect. However, the sampling dis-

[2] For example, β_{YX} is equivalent to a correlation when variables are standardized, and, because $\beta_{MX}\beta_{YM \cdot X} < \beta_{YX}$ when there is no suppression, the standardized indirect effect is bounded by ± 1. Therefore, υ is bounded by 0 and 1.

tribution of the squared standardized indirect effect is not presently known and not easily derived to our knowledge, so the moments of this distribution are not presently available. However, it is possible to derive the bias in the expected value of the sample estimator of υ by application of the known properties of independent, normally distributed random variables that compose $\hat{\upsilon}$.

In order to make use of these properties, it is first necessary to show that υ can be considered the product of squared independent, normally distributed variables. Although the sampling distributions of standardized regression coefficients are not normal, it is the standard assumption that unstandardized regression coefficients are normally distributed across repeated samples when other assumptions are satisfied (i.e., $\hat{B}_{MX} \sim N[B_{MX}, \sigma^2_{MX}]$, $\hat{B}_{YM \cdot X} \sim N[B_{YM \cdot X}, \sigma^2_{YM \cdot X}]$). The second property to consider is that when the model is correctly specified as we assume here, $\hat{B}_{MX}$ and $\hat{B}_{YM \cdot X}$ are independent because (a) the residuals of the M and Y regressions are independent, and (b) $\hat{B}_{YM \cdot X}$ is the relationship between Y and the component of M orthogonal to X, which includes, therefore, the component of X associated with M (i.e., $\hat{B}_{MX}$). The expected value of $\hat{B}_{MX}\hat{B}_{YM \cdot X}$ is equal then to the product of the expected values (i.e., $E[\hat{B}_{MX}\hat{B}_{YM \cdot X}] = E[\hat{B}_{MX}]E[\hat{B}_{YM \cdot X}]$). The third property to consider is that functions of independent random variables are also independent (Casella & Berger, 2002 p. 155), and, thus, $\hat{B}^2_{MX}$ and $\hat{B}^2_{YM \cdot X}$ are independent. Finally, as is common for standardized regression coefficients and standardized indirect effects, we assume that estimates of the variances of X ($\hat{\sigma}^2_X$) and Y ($\hat{\sigma}^2_Y$) are constants and, thus, independent of $\hat{B}_{MX}$, $\hat{B}_{YM \cdot X}$, σ^2_{MX}, and $\sigma^2_{YM \cdot X}$.

Using the definition of the variance of a random variable

$$\text{var}(X) = E[X^2] - (E[X])^2, \tag{17}$$

the expected values of $\hat{B}^2_{MX}$ and $\hat{B}^2_{YM \cdot X}$ can be expressed as

$$\begin{aligned} E[\hat{B}^2_{MX}] &= (E[\hat{B}_{MX}])^2 + \text{var}[\hat{B}_{MX}] = B^2_{MX} + \sigma^2_{MX} \\ E[\hat{B}^2_{YM \cdot X}] &= (E[\hat{B}_{YM \cdot X}])^2 + \text{var}[\hat{B}_{YM \cdot X}] = B^2_{YM \cdot X} + \sigma^2_{YM \cdot X}. \end{aligned} \tag{18}$$

Inserting these results into the definition of the squared standardized indirect effect yields

$$\begin{aligned} &E[\hat{B}^2_{MX}\hat{B}^2_{YM \cdot X}(\hat{\sigma}^2_X/\hat{\sigma}^2_Y)] \\ &= E[\hat{B}^2_{MX}]E[\hat{B}^2_{YM \cdot X}](\hat{\sigma}^2_X/\hat{\sigma}^2_Y) \\ &= (B^2_{MX} + \sigma^2_{MX})(B^2_{YM \cdot X} + \sigma^2_{YM \cdot X})(\hat{\sigma}^2_X/\hat{\sigma}^2_Y) \\ &= (B^2_{MX}B^2_{YM \cdot X} + B^2_{MX}\sigma^2_{YM \cdot X} + B^2_{YM \cdot X}\sigma^2_{MX} + \sigma^2_{MX}\sigma^2_{YM \cdot X})(\hat{\sigma}^2_X/\hat{\sigma}^2_Y). \end{aligned} \tag{19}$$

This result shows that sample estimates of υ using the estimator $\hat{\upsilon}$ are upwardly biased by the factor $B^2_{MX}\sigma^2_{YM \cdot X} + B^2_{YM \cdot X}\sigma^2_{MX} + \sigma^2_{MX}\sigma^2_{YM \cdot X}$.

It can be seen that as N becomes large, variances σ^2_{MX} and $\sigma^2_{YM \cdot X}$ approach zero and, thus, the bias in $\hat{\upsilon}$ approaches zero as well. In other words, the bias in $\hat{\upsilon}$ is negligible in large samples. However, this bias could be problematic at smaller sample sizes, so we derive an estimator of υ that adjusts for this bias.

Because the expected values of $\hat{B}^2_{MX}$ and $\hat{B}^2_{YM \cdot X}$ are the sums of their respective population parameters and sampling variances (see Equation 18), removing the respective sampling variances from $E[\hat{B}^2_{MX}]$ and $E[\hat{B}^2_{YM \cdot X}]$ would yield an expected value of $\hat{\upsilon}$ equal to υ:

$$\begin{aligned} &E[\hat{B}^2_{MX} - \sigma^2_{MX}]E[\hat{B}^2_{YM \cdot X} - \sigma^2_{YM \cdot X}] \\ &= (E[\hat{B}^2_{MX}] - E[\sigma^2_{MX}])(E[\hat{B}^2_{YM \cdot X}] - E[\sigma^2_{YM \cdot X}]) \\ &= (B^2_{MX} + \sigma^2_{MX} - \sigma^2_{MX})(B^2_{YM \cdot X} + \sigma^2_{YM \cdot X} - \sigma^2_{YM \cdot X}) \\ &= B^2_{MX}B^2_{YM \cdot X}. \end{aligned} \tag{20}$$

σ^2_{MX} and $\sigma^2_{YM \cdot X}$ can be replaced with the corresponding unbiased sample estimators $\hat{\sigma}^2_{MX}$ and $\hat{\sigma}^2_{YM \cdot X}$:

$$\begin{aligned} &E[\hat{B}^2_{MX} - \hat{\sigma}^2_{MX}]E[\hat{B}^2_{YM \cdot X} - \hat{\sigma}^2_{YM \cdot X}] \\ &= (E[\hat{B}^2_{MX}] - E[\hat{\sigma}^2_{MX}])(E[\hat{B}^2_{YM \cdot X}] - E[\hat{\sigma}^2_{YM \cdot X}]) \\ &= (B^2_{MX} + \sigma^2_{MX} - \sigma^2_{MX})(B^2_{YM \cdot X} + \sigma^2_{YM \cdot X} - \sigma^2_{YM \cdot X}) \\ &= B^2_{MX}B^2_{YM \cdot X}. \end{aligned} \tag{21}$$

Therefore, the estimator $(\hat{B}^2_{MX} - \sigma^2_{MX})(B^2_{YM \cdot X} - \sigma^2_{YM \cdot X})\ (\hat{\sigma}^2_X / \hat{\sigma}^2_Y)$, to which we hereafter refer as $\tilde{\upsilon}$, is an adjusted estimator that corrects for bias in the expected value of the sample analog of υ.

This adjusted estimator of υ has some important properties that lead us to conclude this estimator is highly promising. Because the parameter estimated by $\tilde{\upsilon}$ in finite samples is υ, it is appropriate to use $\tilde{\upsilon}$ estimates to make inferences about the population υ. As sample size becomes large, the sampling variances of the coefficients $\hat{B}_{MX}$ and $\hat{B}_{YM \cdot X}$ approach zero. Thus, in large samples the adjusted and unadjusted υ are approximately equivalent, and both are unbiased estimators of υ in the limit. Another property is that because the adjusted estimator incorporates the sampling variances of the coefficients that compose the indirect effect, equivalent indirect effect estimates do not necessarily correspond to the same effect size estimate. This means that the factors that affect sampling variances (e.g., sample size, multicollinearity) also affect $\tilde{\upsilon}$. For example, consider indirect effects estimated from two samples with all held equal except coefficient estimates and sampling variances. In the first sample, $\hat{B}_{MX} = 0.6$ ($\hat{\sigma}^2_{MX} = 0.05$), $\hat{B}_{YM \cdot X} = 0.4$ ($\hat{\sigma}^2_{YM \cdot X} = 0.1$), and $\hat{\sigma}^2_X = \hat{\sigma}^2_Y = 1$, resulting in a standardized indirect effect of 0.24. In the second sample, $\hat{B}_{MX} = 0.4$ $\hat{\sigma}^2_{MX} = 0.05$), $\hat{B}_{YM \cdot X} = 0.6$ ($\hat{\sigma}^2_{YM \cdot X} = 0.08$), and $\hat{\sigma}^2_X = \hat{\sigma}^2_Y = 1$, also yielding a standardized indirect effect of 0.24. The unadjusted sample estimates of υ are equivalent ($\hat{\upsilon}_1 = \hat{\upsilon}_2 = 0.058$), but the adjusted estimates differ, with the effect size estimate from the second sample ($\tilde{\upsilon}_2 = 0.025$) larger than the estimate from the first sample ($\tilde{\upsilon}_1 = 0.019$). This is because in Equation 3, in which Y is regressed on M and X, collinearity between M and X is stronger in the first sample than in the second, resulting in a larger standard error for $\hat{B}_{YM \cdot X}$ in the first sample. This suggests that, in general, for equivalent indirect effects and all else being held equal, larger $\hat{B}_{YM \cdot X}$ coefficients have correspondingly larger effect sizes, which is consistent with previous research examining the behavior of different specifications for indirect effects (Beasley, 2014; Fritz & MacKinnon, 2007; Fritz, Taylor, & MacKinnon, 2012; Hoyle & Kenny, 1999).

Simulation Studies

Monte Carlo simulations are often employed to examine the finite sample properties of statistical estimators. Employing this type of simulation allows for the behavior of estimators to be studied under various conditions (e.g., small N, nonnormality, violation of assumptions, model misspecification). Additionally, Monte Carlo simulations are useful for understanding various properties of estimators with asymptotic distributions that are complex or unknown. For simple three-variable mediation models, υ is the square of the standardized indirect effect, the sampling distribution of which is complex. The distribution of the square of the standardized indirect effect is not known to have been derived. Because the distribution of this effect size is unknown, we have used Monte Carlo simulations to facilitate our understanding of the measure. In addition, because Monte Carlo simulations can be used to validate theoretical results, we have used Monte Carlo simulations to determine whether the properties of the adjusted and unadjusted estimators derived in the previous section were consistent with the observed properties of the estimators.

The Monte Carlo simulation study was designed to examine the finite sample properties of adjusted and unadjusted estimators of υ in a simple three-variable mediation model from populations with varying but known magnitudes of indirect effects. For point estimators, of interest were absolute bias, percent relative bias, mean square error (MSE), degree of consistency, and degree of efficiency. 95% CIs were constructed using a percentile bootstrap procedure, in which of interest were CI widths, coverage, and the proportions of estimates above and below the 95% CI limits.

Bias was defined as the difference between the expectation of the sample estimator ($\bar{\hat{\theta}}$) and the parameter that it estimates (θ):

$$bias(\hat{\theta}) = \bar{\hat{\theta}} - \theta. \quad (22)$$

All else being equal, lower bias for an estimator indicates that the expected value is on average closer to the parameter compared with a competing estimator; an unbiased estimator has a bias of zero. Because υ quantifies explained variance, we anticipate that the unadjusted $\hat{\upsilon}$ would have bias similar to the bias inherent in unadjusted estimators of R^2. Note that this is not a limitation of our effect size per se, but rather a limitation of the current unadjusted estimates of R^2, on which $\hat{\upsilon}$ are based. Therefore, it was hypothesized that $\hat{\upsilon}$ would be upwardly biased, particularly for the smallest N and smallest effects, and that bias would decrease at larger N and for larger effects. However, it would be expected that bias of $\tilde{\upsilon}$ would be acceptable for all N and effect sizes considered.

Percent relative bias was defined as the ratio of the bias to the parameter value:

$$bias_{rel}(\hat{\theta}) = \frac{\bar{\hat{\theta}} - \theta}{\theta} \times 100. \quad (23)$$

Like bias, with all else being equal, lower relative bias for an estimator means that the expected value of the sample estimator is on average closer to the population parameter compared with a competing estimator; an unbiased estimator has a relative bias of zero. However, relative bias is on a different scale than bias; whereas the measures convey similar information, relative bias provides results that may be more interpretable because it is scaled relative to the size of the population value being estimated. We regard both measures as useful. We considered percent relative bias <5% to be acceptable (Boomsma, 2013).

MSE was defined as the sum of the variance of sample estimates and squared bias,

$$MSE = [bias(\hat{\theta})]^2 + \mathrm{var}(\hat{\theta}). \quad (24)$$

MSE is a measure of the bias and variability (the inverse of precision) of a sample estimator and serves as a measure of accuracy; a smaller (lower) MSE indicates a more accurate estimator. When a sample estimator is unbiased, MSE is equivalent to the variance of that estimator. Although we hypothesized that $\hat{\upsilon}$ would be positively biased, we also hypothesized that the bias would decrease as sample size increased. In addition, the variance would be expected to decrease with increasing N. Therefore, we hypothesized that MSE would decrease as N increased for $\hat{\upsilon}$. We hypothesized that $\tilde{\upsilon}$ would be unbiased for all conditions, so changes in N would not be associated with changes in bias for this estimator. However, we expected that the variance of $\tilde{\upsilon}$ would decrease with increasing N, and, therefore, MSE would decrease with increasing N.

Consistency of the estimators was evaluated by conducting the simulations across multiple values of N. As N is increased, the statistics were expected to converge to their population values (i.e., bias and variance would decrease as N increased). Moreover, MSE and CI width were expected to decrease as N increased. Efficiency of the estimators was evaluated by comparing the CI widths, in which smaller CI widths indicate more efficient estimators. Comparisons were conducted taking the ratio of CI widths (i.e., relative efficiency). Relative efficiencies greater than 1 indicated that the estimator in the numerator was more efficient, and relative efficiencies less than 1 indicated that the estimator in the denominator was more efficient. It was unclear which estimator, $\hat{\upsilon}$ or $\tilde{\upsilon}$, would be more efficient, so this was addressed empirically.

Coverage was defined as the proportion of samples in which the population parameter was contained within the 95% CI. We considered acceptable coverage to be between 92.5% and 97.5% (Bradley, 1978), and hypothesized that coverage would approach a nominal level of 95% as N increased. If the population parameter was not contained within a 95% CI, it was recorded whether the parameter was above the upper CI limit or below the lower CI limit. A proper 95% CI with equal error rates in each tail requires 2.5% of samples outside of the CI to be above the upper confidence limit and 2.5% below the lower confidence limit. It was hypothesized that the proportion of these misses to the left and right of the CI would be equal.

The three-variable mediation model in Figure 1 was the generating model for the simulation study. X, M, and Y were standardized normal variates with means of zero and variances of 1. Regression coefficients composing the indirect effect were varied among .15, .39, and .59, corresponding to small, medium, and large effect sizes (Cohen, 1988). The direct effect was varied among 0, .15, and .39. We did not consider models for which the population indirect effect was zero because the bootstrap estimates have been shown to be unreliable when the parameter of interest is on the boundary of the parameter space (Andrews, 2000; Chernick, 2008). 10,000 replication samples were created using the `mvrnorm()` function from the MASS package (Venables & Ripley, 2002) in R (Version 3.1.2; R Development Core Team, 2014). 10,000 replications were sufficient to achieve reasonably accurate estimates of bias and empirical 95% CI coverage.

For each of the 10,000 replications, 1,000 bootstrap resamples were created by sampling cases with replacement, and 95% CIs were constructed from the 2.5th and 97.5th percentiles of the empirical distribution of estimates (i.e., percentile bootstrap CIs). Values for the upper and lower percentiles, CI width, and whether or not the population parameter was within the CI were recorded for each replication. When the population parameter was outside the 95% CI, we also recorded whether the parameter was above the upper CI limit or below the lower limit. $\tilde{\upsilon}$ and $\hat{\upsilon}$ were also estimated for each of the 10,000 replications, and means and variances of each statistic were computed across replications. This procedure was performed for $N = 50$, 100, 250, and 500, which are representative of sample sizes commonly used in mediation models, based on our experience.

Results of Monte Carlo Simulation

Results from the simulation study can be found in Tables 1 through 5. The hypothesis that $\hat{\upsilon}$ would be upwardly biased, particularly at smaller N and for smaller effect sizes, and less biased as N and effect size increased, was supported by the simulation results (see Table 1). The direction of bias for $\hat{\upsilon}$ was positive in all conditions. This is not surprising considering it was demonstrated analytically that $\hat{\upsilon}$ is a positively biased estimator of υ. The largest values of percent relative bias (264.3%, 279.779%, and 216.242%) occurred at the smallest N considered in the simulation ($N = 50$), and also for the smallest effects ($\beta^2_{MX} = \beta^2_{YM \cdot X} = .15$). In addition, for smaller effect sizes, percent relative bias >5% was evident even at the largest N considered ($N = 500$). Finally, bias decreased as N increased, supporting the hypothesis that $\hat{\upsilon}$ is a consistent estimator.

Simulation results of percent relative bias for $\tilde{\upsilon}$ can also be found in Table 1. The hypothesis that bias of $\tilde{\upsilon}$ would be acceptable across simulation conditions was largely supported by simulation results. Overall, percent relative biases for $\tilde{\upsilon}$ were of much smaller magnitude than for $\hat{\upsilon}$. For the conditions in which bias was greatest for the $\hat{\upsilon}$ ($N = 50$, $\beta^2_{MX} = \beta^2_{YM \cdot X} = .15$), the relative biases of $\tilde{\upsilon}$ were −6.783%, 11.104%, and −14.648%. In total, the largest relative bias across all N and effect sizes considered for $\tilde{\upsilon}$ was −14.648%. In terms of raw values, the υ value for this parameter combination was 0.000506, and the average $\tilde{\upsilon}$ was 0.000432. At $N = 100$, only four parameter combinations had percent relative bias >5% (−5.245%, −5.987, −8.052, and −5.749%), and no combination had percent relative bias >5% at $N = 250$ and $N = 500$. Although As expected, bias was largest at the smallest effect size considered, and there did not appear to be a systematic relationship between population effect size and bias for $\tilde{\upsilon}$ as there was for $\hat{\upsilon}$. Finally, as with $\hat{\upsilon}$, bias decreased as N increased, supporting the hypothesis that $\tilde{\upsilon}$ is a consistent estimator.

Simulation results of MSE can be found in Table 2 and CI width in Table 3. It was shown that as N increased, MSE values approached zero and CI widths narrowed for both measures, supporting the hypothesis that overall accuracy of the measures would increase with increasing N. Given that bias for $\hat{\upsilon}$ was consistently larger than for $\tilde{\upsilon}$, it was not surprising that $\tilde{\upsilon}$ consistently exhibited smaller values of MSE across all parameter combinations. Results also show that average CI widths for both estimators were comparable across conditions, with relative efficiencies of $\tilde{\upsilon}$ to $\hat{\upsilon}$ ranging from approximately 0.9 to 1.0. Interestingly, although of comparable magnitude, CI widths for $\tilde{\upsilon}$ were consistently narrower than CI widths for $\hat{\upsilon}$ for the vast majority of simulation conditions, suggesting that $\tilde{\upsilon}$ is a more efficient estimator.

Table 1
Percent Relative Bias of the Unadjusted $\hat{\upsilon}$ and Adjusted $\tilde{\upsilon}$

			$\hat{\upsilon}$				$\tilde{\upsilon}$			
β_{MX}	$\beta_{YM \cdot X}$	$\beta_{YX \cdot M}$	$N = 50$	$N = 100$	$N = 250$	$N = 500$	$N = 50$	$N = 100$	$N = 250$	$N = 500$
.15	.15	0	264.300	103.912	36.558	17.599	−6.783	−4.971	−1.752	−.816
		.15	279.779	108.010	36.409	17.364	11.104	−.243	−1.349	−.719
		.39	216.242	95.292	34.024	17.231	−14.648	−2.110	−.883	.448
	.39	0	107.258	50.284	20.019	9.792	−2.715	−1.461	−.086	−.151
		.15	98.362	47.324	19.148	9.894	−8.765	−3.886	−.879	.023
		.39	99.667	44.701	17.112	8.523	−4.769	−5.245	−2.343	−1.101
	.59	0	86.547	43.681	19.739	9.003	−7.410	−2.579	1.446	−.116
		.15	83.967	40.001	16.593	7.903	−9.730	−5.987	−1.603	−1.173
		.39	79.060	37.402	14.493	8.039	−14.169	−8.052	−3.489	−.940
.39	.15	0	129.587	60.146	21.388	12.492	−4.973	−1.600	−1.991	.907
		.15	121.046	53.723	23.100	10.891	−8.062	−5.749	.443	−.211
		.39	109.103	49.998	19.106	8.354	−.768	−.399	−.054	−1.049
	.39	0	23.214	10.601	4.088	1.869	−3.633	−2.143	−.867	−.583
		.15	23.481	9.619	4.188	2.459	−2.184	−2.553	−.554	.108
		.39	16.731	8.220	2.564	1.673	−5.122	−2.270	−1.514	−.354
	.59	0	13.385	6.765	2.528	.975	−3.074	−1.244	−.623	−.590
		.15	12.110	5.091	1.975	.529	−3.655	−2.554	−1.042	−.971
		.39	7.811	3.715	1.845	.677	−6.142	−3.090	−.842	−.660
.59	.15	0	152.448	70.384	26.387	12.584	−2.524	−1.978	−1.566	−1.261
		.15	143.485	67.294	25.725	13.881	−2.303	−1.822	−.950	.672
		.39	112.354	55.489	23.110	11.426	−6.754	−.755	1.321	.616
	.39	0	20.691	8.311	3.531	1.639	−2.365	−2.624	−.721	−.467
		.15	17.812	8.938	3.439	1.149	−3.076	−1.054	−.429	−.767
		.39	11.038	5.780	2.071	1.317	−4.340	−1.600	−.808	−.112
	.59	0	6.690	2.921	1.279	.388	−3.508	−1.997	−.647	−.569
		.15	5.766	3.086	.770	.555	−3.207	−1.246	−.925	−.288
		.39	2.240	.900	.395	.133	−3.862	−2.064	−.770	−.446

Table 2
Mean Square Error for the Unadjusted $\hat{\upsilon}$ and Adjusted $\tilde{\upsilon}$

			$\hat{\upsilon}$				$\tilde{\upsilon}$			
β_{MX}	$\beta_{YM \cdot X}$	$\beta_{YX \cdot M}$	$N = 50$	$N = 100$	$N = 250$	$N = 500$	$N = 50$	$N = 100$	$N = 250$	$N = 500$
.15	.15	0	.00002	<.00001	<.00001	<.00001	.00001	<.00001	<.00001	<.00001
		.15	.00002	<.00001	<.00001	<.00001	.00001	<.00001	<.00001	<.00001
		.39	.00001	<.00001	<.00001	<.00001	.00001	<.00001	<.00001	<.00001
	.39	0	.00012	.00004	.00001	.00001	.00008	.00003	.00001	<.00001
		.15	.00011	.00004	.00001	<.00001	.00008	.00003	.00001	<.00001
		.39	.00010	.00003	.00001	<.00001	.00007	.00003	.00001	<.00001
	.59	0	.00038	.00015	.00005	.00002	.00031	.00013	.00005	.00002
		.15	.00036	.00014	.00005	.00002	.00030	.00012	.00004	.00002
		.39	.00030	.00013	.00004	.00002	.00026	.00012	.00004	.00002
.39	.15	0	.00017	.00005	.00001	.00001	.00012	.00004	.00001	.00001
		.15	.00015	.00005	.00001	.00001	.00010	.00004	.00001	.00001
		.39	.00013	.00004	.00001	<.00001	.00009	.00003	.00001	<.00001
	.39	0	.00081	.00031	.00011	.00005	.00066	.00028	.00010	.00005
		.15	.00077	.00028	.00010	.00005	.00063	.00025	.00009	.00005
		.39	.00054	.00022	.00008	.00004	.00045	.00020	.00008	.00004
	.59	0	.00202	.00087	.00031	.00016	.00180	.00082	.00031	.00015
		.15	.00182	.00078	.00028	.00014	.00165	.00075	.00028	.00014
		.39	.00128	.00059	.00023	.00011	.00123	.00058	.00023	.00011
.59	.15	0	.00097	.00028	.00008	.00004	.00075	.00024	.00008	.00003
		.15	.00086	.00028	.00008	.00004	.00066	.00024	.00007	.00003
		.39	.00057	.00020	.00006	.00003	.00044	.00017	.00006	.00003
	.39	0	.00308	.00125	.00046	.00022	.00271	.00117	.00045	.00022
		.15	.00263	.00115	.00041	.00020	.00233	.00107	.00039	.00019
		.39	.00177	.00079	.00029	.00014	.00160	.00075	.00028	.00014
	.59	0	.00565	.00254	.00095	.00046	.00523	.00245	.00093	.00045
		.15	.00464	.00209	.00078	.00039	.00433	.00201	.00077	.00038
		.39	.00249	.00119	.00045	.00023	.00243	.00118	.00044	.00023

The hypotheses that coverage would reach the nominal 95% level as *N* increased, and that the proportions of misses to the left and right of the 95% CI would be balanced, were partially supported by simulation results. Results for coverage and proportions of misses to the left and right for $\hat{\upsilon}$ can be found in Table 4, and results for $\tilde{\upsilon}$ can be found in Table 5. As with bias, satisfactory coverage was achieved with larger effect sizes and at larger *N* for both measures. For both estimators, nominal coverage was satisfactory at the smallest *N* for some parameter combinations, but for most combinations, coverage achieved approximately nominal levels as *N* increased. For both $\hat{\upsilon}$ and $\tilde{\upsilon}$, satisfactory coverage was achieved for all parameter combinations at $N = 250$. For $\hat{\upsilon}$, the proportions of misses to the left and right were approximately balanced when $N = 250$ and $N = 500$. For $\tilde{\upsilon}$, although coverage was satisfactory, there appeared to be a consistent imbalance in the proportion of misses, such that it was more likely for misses to be greater than the upper confidence limit, even at the largest *N* considered.

Software Implementation

For simple three-variable mediation models, $\hat{\upsilon}$ is calculated by simply standardizing the squared indirect effect, and $\tilde{\upsilon}$ by standardizing the product of the squared coefficients after subtracting the respective estimated coefficient variances. The estimated coefficient variances can typically be found in the estimated asymptotic covariance matrix. However, obtaining bootstrap CIs for the estimators can be challenging. To facilitate the use of υ for researchers, we have developed a set of R functions and incorporated them into the MBESS (K. Kelley, 2007a, 2007b, 2017) R (R Development Core Team, 2010) package. The function `mediation()` has been updated to obtain unadjusted and adjusted estimates of υ, with either percentile, bias-corrected, or bias-corrected and accelerated bootstrap CIs. The `mediation()` function accepts either raw data or summary statistics as input. Documentation for the functions is contained within the MBESS package.

Empirical Example

Here we present an empirical example to facilitate interpretation and implementation of the unadjusted and adjusted υ estimators. We use results from an RCT examining the effects of cognitive–behavioral therapy (CBT) for social anxiety disorder (SAD) by Goldin et al. (2012). The purpose of the study was to examine whether the effect of CBT on reducing symptoms of social anxiety was mediated by changes in a certain type of self-belief called *cognitive reappraisal self-efficacy* (CR-SE). The investigators randomized 75 individuals who satisfied inclusionary criteria and met diagnostic criteria for SAD to treatment (CBT; $N = 38$) and wait-list control (WL; $N = 37$) conditions. Participants in the treatment condition received weekly sessions of CBT administered by trained clinical psychologists for 4 months. By the end of treatment, six participants had dropped out from the CBT condition, and five from the WL condition. The mediator variable CR-SE was measured using the self-efficacy subscale of the Emotion Regulation Questionnaire. Scores on the eight-item subscale were summed, such that higher values indicated greater perceived

Table 3
95% Confidence Interval Widths of the Unadjusted $\hat{\upsilon}$ and Adjusted $\tilde{\upsilon}$

			$\hat{\upsilon}$				$\tilde{\upsilon}$			
β_{MX}	$\beta_{YM \cdot X}$	$\beta_{YX \cdot M}$	$N = 50$	$N = 100$	$N = 250$	$N = 500$	$N = 50$	$N = 100$	$N = 250$	$N = 500$
.15	.15	0	.0208	.0082	.0034	.0020	.0189	.0075	.0031	.0019
		.15	.0205	.0082	.0033	.0020	.0187	.0075	.0031	.0019
		.39	.0176	.0073	.0031	.0019	.0160	.0067	.0029	.0018
	.39	0	.0459	.0244	.0129	.0087	.0432	.0233	.0125	.0085
		.15	.0437	.0236	.0126	.0085	.0413	.0227	.0123	.0084
		.39	.0409	.0220	.0120	.0082	.0390	.0213	.0118	.0081
	.59	0	.0780	.0461	.0266	.0183	.0763	.0455	.0264	.0183
		.15	.0749	.0442	.0256	.0179	.0738	.0438	.0255	.0178
		.39	.0698	.0412	.0242	.0172	.0699	.0413	.0242	.0172
.39	.15	0	.0568	.0282	.0141	.0095	.0545	.0272	.0137	.0093
		.15	.0540	.0270	.0140	.0092	.0517	.0260	.0136	.0091
		.39	.0474	.0244	.0126	.0084	.0453	.0234	.0123	.0083
	.39	0	.1079	.0679	.0401	.0277	.1018	.0651	.0393	.0274
		.15	.1035	.0649	.0387	.0268	.0978	.0623	.0379	.0266
		.39	.0897	.0581	.0347	.0242	.0851	.0560	.0342	.0240
	.59	0	.1679	.1128	.0695	.0486	.1617	.1104	.0688	.0483
		.15	.1566	.1056	.0658	.0462	.1517	.1037	.0653	.0460
		.39	.1323	.0925	.0588	.0414	.1308	.0919	.0586	.0414
.59	.15	0	.1286	.0662	.0340	.0229	.1283	.0659	.0338	.0228
		.15	.1215	.0637	.0331	.0225	.1211	.0634	.0329	.0223
		.39	.1016	.0550	.0298	.0203	.1010	.0546	.0295	.0202
	.39	0	.2112	.1354	.0826	.0576	.2053	.1326	.0818	.0573
		.15	.1963	.1289	.0781	.0544	.1906	.1263	.0773	.0542
		.39	.1611	.1080	.0659	.0462	.1561	.1059	.0653	.0460
	.59	0	.2896	.1941	.1197	.0839	.2814	.1909	.1188	.0836
		.15	.2623	.1777	.1091	.0767	.2555	.1751	.1084	.0765
		.39	.1939	.1339	.0835	.0587	.1914	.1329	.0832	.0586

capability to use cognitive reappraisals. The outcome variable was measured using the Liebowitz Social Anxiety Scale (LSAS). Scores on the 24-item scale were summed, such that higher values indicated greater symptoms of social anxiety. CR-SE and SA symptoms were assessed at pre- and posttreatment. Following the original analysis, the mediator variable was defined as the change in CR-SE (ΔCR-SE) from pre- to posttreatment, and outcome was the LSAS score at posttreatment. The analytic sample consisted of participants that completed pre- and posttreatment assessments of both measures in the CBT ($N = 32$) and WL ($N = 32$) conditions.

We did not have access to the raw data from this study. Because the percentile bootstrap requires resampling individual cases, we simulated data using the descriptive statistics and results reported in the original manuscript. Means and standard deviations for simulated observations of ΔCR-SE (CBT = 15.55, $SD = 12.03$; WL = −2.12, $SD = 11.84$) and LSAS (CBT = 53.45, $SD = 241$; WL = 70.01, $SD = 18.94$) were comparable with those in the original study. It is important to note that it is assumed that the mediation model is correctly specified, and the model presented here for demonstration purposes is an overly simplistic representation of this potential causal process. A more rigorous investigation could control for pretreatment LSAS, and assess potential confounding of the ΔCR-SE/LSAS relationship using sensitivity analysis (Imai, Keele, & Yamamoto, 2010).

Table 6 presents results for $\hat{\upsilon}$ and $\tilde{\upsilon}$, as well as basic results for raw regression coefficients. Because the standardized indirect effect and *MedES* are also viable effect sizes for this design, we include estimates and CIs for these measures for comparison. The raw indirect effect of CBT on LSAS through ΔCR-SE was −11.792 (95% percentile bootstrap CI [−20.675, −3.122]), which was significantly different from zero. The corresponding partially standardized indirect effect size estimate was −0.508 (95% percentile bootstrap CI [−0.89, −0.14]), and the corresponding *MedES* effect size estimate was 0.287 (95% percentile bootstrap CI [0.065, 0.517]). For the υ estimators, the $\hat{\upsilon}$ effect size estimate was 0.065 (95% percentile bootstrap CI [0.006, 0.201]), meaning that the variance explained indirectly in LSAS by CBT through ΔCR-SE in this sample was 0.065. Assuming Cohen's effect size standards are appropriate for this study, this result would be considered a small to medium effect size. The $\tilde{\upsilon}$ effect size estimate was 0.057 (95% percentile bootstrap CI [−0.005, 0.188]). $\tilde{\upsilon}$ is interpreted as the estimated variance in LSAS explained by CBT through ΔCR-SE in the population. By Cohen's standards, this would also be considered a small to medium effect size.

Discussion

The goal of this research was to develop an effect size measure that addressed the limitations of existing effect sizes for mediation analysis. A review of currently available mediation effect sizes revealed that many had deficiencies that limited their use. For the mediation models considered in this study, υ was shown to be an appropriate and interpretable measure of effect size. It was demonstrated that υ represents the variance in the outcome explained jointly by the predictor and mediator. υ builds upon the explained variance framework proposed by Fairchild et al. (2009) and MacKinnon (2008) by using decomposition techniques from path analysis to account for spurious correlation unassociated with the

Table 4
Percent Confidence Interval Coverage, and Proportions of Misses to the Left and Right of the Confidence Interval of the Unadjusted $\hat{\upsilon}$

			N = 50			N = 100			N = 250			N = 500		
β_{MX}	$\beta_{YM \cdot X}$	$\beta_{YX \cdot M}$	Cov	ML	MH	Cov	ML	MH	Cov	ML	MH	Cov	ML	MH
.15	.15	0	99.2	.8	.0	99.0	1.0	.1	95.7	1.3	3.0	94.7	1.6	3.7
		.15	99.0	1.0	.0	99.1	.9	.0	95.6	1.2	3.2	94.6	1.7	3.7
		.39	99.3	.7	.0	98.7	1.1	.2	95.3	1.3	3.5	94.4	1.6	4.0
	.39	0	98.3	1.5	.2	97.6	1.8	.6	95.4	2.1	2.5	94.7	2.4	2.9
		.15	98.2	1.6	.2	97.7	1.8	.6	95.2	2.2	2.6	95.0	2.1	2.9
		.39	98.2	1.7	.2	97.6	2.0	.4	95.6	2.1	2.2	94.6	2.5	2.9
	.59	0	97.7	2.3	.0	97.4	2.5	.1	95.1	2.7	2.2	95.0	2.5	2.6
		.15	97.5	2.4	.1	97.7	2.2	.1	95.8	2.3	1.9	94.6	2.7	2.8
		.39	97.6	2.4	.0	97.4	2.6	.0	95.4	2.7	1.8	94.0	3.0	3.0
.39	.15	0	98.0	1.9	.2	97.4	2.1	.5	96.2	2.1	1.8	94.6	2.6	2.8
		.15	98.3	1.5	.2	97.7	1.9	.4	96.0	2.4	1.6	94.9	2.3	2.8
		.39	97.8	1.8	.3	97.3	2.1	.6	95.0	2.4	2.7	94.4	2.3	3.3
	.39	0	94.2	1.8	4.0	94.5	1.8	3.7	94.2	2.3	3.6	94.9	2.1	2.9
		.15	94.2	1.8	4.1	94.7	1.8	3.5	94.9	2.0	3.1	94.7	2.3	3.0
		.39	94.3	1.6	4.1	94.4	1.9	3.7	94.6	2.1	3.4	94.9	2.3	2.8
	.59	0	93.3	2.3	4.4	94.2	2.2	3.6	94.5	2.3	3.2	94.8	2.2	3.0
		.15	93.5	2.3	4.2	93.8	2.4	3.8	94.4	2.5	3.1	95.0	2.2	2.9
		.39	94.1	2.4	3.5	94.5	2.2	3.2	94.3	2.6	3.1	94.6	2.5	2.9
.59	.15	0	97.1	2.9	.0	97.5	2.5	.0	96.8	2.7	.5	95.2	2.3	2.5
		.15	97.0	3.0	.0	97.0	3.0	.0	96.7	2.6	.7	95.0	2.7	2.2
		.39	97.3	2.7	.0	97.3	2.7	.1	96.2	2.6	1.2	94.5	2.6	2.8
	.39	0	94.9	2.3	2.8	94.3	2.4	3.3	94.8	2.4	2.8	94.5	2.6	2.8
		.15	94.7	2.3	3.0	94.3	2.5	3.2	94.5	2.5	3.1	94.2	2.6	3.2
		.39	94.2	2.1	3.6	94.3	2.5	3.2	94.8	2.4	2.8	94.5	2.7	2.8
	.59	0	94.0	2.2	3.8	94.2	2.5	3.3	94.6	2.3	3.1	94.8	2.2	3.0
		.15	93.9	2.4	3.7	94.0	2.7	3.3	94.9	2.1	3.0	94.8	2.3	2.9
		.39	94.4	2.2	3.5	94.4	2.3	3.2	95.2	2.0	2.8	94.5	2.4	3.1

Note. Cov = % coverage; MH = % misses higher than upper confidence limit; ML = % misses below lower confidence limit.

indirect effect. Accounting for this spurious effect corrects a contradictory result observed in R^2_{med}, in which an indirect effect of zero can return a nonzero effect size. It was also shown that υ is equivalent to the squared standardized indirect effect, and represents the variance in an outcome explained indirectly by a predictor through a mediator.

This measure has many desirable properties as an effect size for the indirect effect in mediation analysis. First, υ has an interpretable scale as a R^2 measure of effect size, and can appropriately be compared with benchmarks for R^2 detailed in Cohen (1988). Because the measure is a function of standardized regression coefficients, υ is also standardized; that is, the measure does not depend on the scales of the predictor, mediator, or outcome, and is invariant under linear transformations of the variables. υ is a monotonically increasing function in absolute value of the standardized indirect effect. This means that, with all else held constant, a larger effect size directly corresponds to a larger magnitude of the indirect effect. Second, Monte Carlo simulation demonstrated that CIs can be constructed for the estimators for $\hat{\upsilon}$ and $\tilde{\upsilon}$ using a bootstrap procedure, and that the percentile bootstrap CIs show satisfactory coverage for the majority of parameter combinations considered. Third, we showed that bias in the unadjusted estimator $\hat{\upsilon}$ was acceptable for many parameter combinations at N = 250 and N = 500 and presumably larger sample sizes. Importantly, however, bias in the adjusted estimator $\tilde{\upsilon}$ was acceptable for the vast majority of simulation conditions. Fourth, $\hat{\upsilon}$ and $\tilde{\upsilon}$ are consistent estimators of υ. Fifth, $\hat{\upsilon}$ and $\tilde{\upsilon}$ were comparable in terms of CI widths, with $\tilde{\upsilon}$ estimates exhibiting a slight but consistent advantage in precision. For these reasons, we believe $\tilde{\upsilon}$ has substantial advantages over $\hat{\upsilon}$ as a sample estimator of υ, especially for studies with smaller N and for smaller effects. We therefore are encouraged enough by our analytic justifications and Monte Carlo simulation results to widely recommend use of $\tilde{\upsilon}$ and believe that it will help researchers in a wide variety of areas to better communicate the size of mediation effects.

Comparisons With Other Effect Size Measures for Indirect Effects

It should be noted that the standardized indirect effect shares many of the desirable properties of υ. The standardized indirect effect is interpretable in terms similar to the interpretation of a standardized regression coefficient, invariant to appropriate linear transformations (e.g., linear transformation of a categorical predictor would not be considered appropriate), and independent of sample size. In this regard, for the indirect effect from a three-variable mediation model, defining the effect size parameter of interest as either the standardized indirect effect or υ is a matter of the desired interpretation. Because in the population, with all else held constant, υ is a monotonic function in absolute value of the standardized indirect effect, the effect size measures are essentially conveying the same information. This is analogous to the relationship between a standardized regression coefficient and R^2 for a simple regression model in the population. However, the information conveyed by the sample estimator of the standardized indirect effect and $\tilde{\upsilon}$ would not be redundant. The advantage of reporting $\tilde{\upsilon}$

Table 5
Percent Confidence Interval Coverage, and Proportions of Misses to the Left and Right of the Confidence Interval of the Adjusted $\tilde{\upsilon}$

			N = 50			N = 100			N = 250			N = 500		
β_{MX}	$\beta_{YM\cdot X}$	$\beta_{YX\cdot M}$	Cov	ML	MH	Cov	ML	MH	Cov	ML	MH	Cov	ML	MH
.15	.15	0	99.4	.1	.6	96.2	.1	3.7	92.7	.4	6.9	93.3	.8	5.9
		.15	99.3	.1	.6	95.8	.1	4.0	92.7	.5	6.8	93.3	1.0	5.7
		.39	99.2	.0	.8	95.4	.2	4.4	92.5	.6	6.8	93.0	.9	6.1
	.39	0	97.9	.3	1.8	97.0	.6	2.4	93.7	1.4	4.9	94.2	1.6	4.1
		.15	97.9	.4	1.7	97.2	.6	2.1	94.2	1.1	4.7	94.3	1.5	4.2
		.39	98.4	.5	1.1	97.4	.8	1.8	94.5	1.2	4.4	94.0	1.8	4.2
	.59	0	99.0	.7	.3	98.0	1.3	.7	94.2	1.8	4.0	94.7	1.7	3.6
		.15	99.0	.7	.3	98.4	1.1	.5	94.4	1.5	4.1	94.5	1.9	3.7
		.39	99.0	.9	.1	98.5	1.3	.2	94.2	1.6	4.2	93.7	2.2	4.1
.39	.15	0	98.0	.4	1.5	97.1	.8	2.1	94.5	1.3	4.2	94.0	1.9	4.1
		.15	98.0	.2	1.7	97.4	.8	1.8	95.0	1.3	3.7	94.4	1.5	4.1
		.39	97.7	.5	1.8	96.7	.9	2.4	94.0	1.3	4.7	93.7	1.6	4.7
	.39	0	91.8	.8	7.4	93.4	1.3	5.3	93.8	1.6	4.6	94.8	1.7	3.5
		.15	92.0	.8	7.2	93.6	1.0	5.4	94.3	1.6	4.1	94.5	2.0	3.6
		.39	92.2	.7	7.0	93.5	1.2	5.3	94.3	1.6	4.1	94.7	1.9	3.4
	.59	0	92.2	1.4	6.5	93.7	1.4	4.9	94.5	1.7	3.8	94.8	1.8	3.4
		.15	92.3	1.4	6.3	93.2	1.7	5.1	94.2	2.0	3.8	94.7	1.9	3.3
		.39	93.1	1.4	5.5	94.1	1.5	4.4	94.2	1.9	3.8	94.3	2.3	3.4
.59	.15	0	98.9	1.0	.1	98.9	1.0	.2	96.4	1.5	2.1	94.6	1.5	3.9
		.15	99.0	.9	.2	98.5	1.3	.2	96.2	1.5	2.3	94.7	1.8	3.6
		.39	98.8	.8	.3	98.2	1.2	.6	95.2	1.4	3.4	94.3	1.8	3.8
	.39	0	93.6	1.2	5.2	93.5	1.5	4.9	94.5	2.0	3.5	94.4	2.2	3.4
		.15	93.5	1.3	5.2	93.7	1.8	4.6	94.3	1.9	3.8	94.2	2.1	3.6
		.39	92.7	1.4	5.9	94.0	1.7	4.3	94.6	2.0	3.4	94.4	2.4	3.2
	.59	0	93.3	1.3	5.4	93.9	1.9	4.3	94.3	2.0	3.8	94.6	2.0	3.4
		.15	93.5	1.5	5.0	93.8	2.0	4.2	94.8	1.8	3.5	94.8	2.0	3.1
		.39	93.6	1.6	4.9	94.1	1.9	4.1	94.9	1.8	3.4	94.4	2.1	3.4

Note. Cov = % coverage; MH = % misses higher than upper confidence limit; ML = % misses below lower confidence limit.

is therefore analogous to that of reporting the adjusted R^2 for a simple linear regression model, as the R^2 adjustment penalizes estimates for imprecision due to sample size and collinearity. Alternatively, an advantage of the standardized indirect effect is that the CI may include zero, making the measure more useful for NHST purposes. However, although a useful property, utility in an NHST framework is generally not a necessary property for effect size measures; the purpose of an effect size is to quantify the magnitude and precision of an effect of interest on a more meaningful or interpretable metric, regardless of statistical significance. Overall, because of the complexity inherent in even the simplest mediation model, it is unlikely that a single effect size measure can adequately capture all the ways in which indirect effects may differ, and therefore we would recommend reporting both the standardized indirect effect and $\tilde{\upsilon}$.

Table 6
Estimates, Standard Errors, and Confidence Limits for the Mediation Model in the Goldin et al. (2008) Empirical Example

				95% CI	
Statistic	Estimate	*SE*	*p*	LCL	UCL
Δ CR-SE					
Intercept	6.72	1.49	<.001		
CBT	17.67	2.98	<.001		
LSAS					
Intercept	66.21	2.95	<.001		
Δ CR-SE	−.67	.22	<.05		
CBT	−4.77	6.42	.46		
$\hat{B}_{MX}\hat{B}_{YM\cdot X}$	−11.790			−20.675	−3.122
$\beta_{MX}\beta_{YM\cdot X}$	−.256			−.449	−.070
MedES	.287			.065	.517
$\hat{\upsilon}$	.065			.006	.201
$\tilde{\upsilon}$	.057			−.005	.188

Note. ΔCR-SE = pre- to posttreatment change in cognitive reappraisal self-efficacy; CBT = cognitive behavioral therapy; LSAS = Liebowitz Social Anxiety Scale; *SE* = standard error; LCL = lower 95% percentile bootstrap CI limit; UCL = upper 95% percentile bootstrap CI limit; $\hat{\upsilon}$ = unadjusted effect size; $\tilde{\upsilon}$ = adjusted effect size.

Given the importance of binary predictors in psychological research, particularly for coding group assignment in RCTs, and because υ is applicable as an effect size for these indirect effects, it is also of interest to compare the properties of υ with those of *MedES*. It is important to note that *MedES* was developed in the MacArthur framework for mediation analysis, which requires modeling the interaction of the predictor and mediator, and υ has yet to be extended to this class of mediation models. However, from a structural equation modeling perspective, *MedES* can be reexpressed omitting the interaction so the measures can be directly compared. In the context of a binary predictor, *MedES* translates the indirect effect into a difference in success rate differences (SRDs), which gives the effect size an interpretable scale with benchmarks for small, medium, and large obtainable from benchmarks for Cohen's *d* (Kraemer & Kupfer, 2006). In addition, computation of *MedES* requires converting the total and direct effects into Cohen's *d*, meaning that *MedES* is partially standardized in the sense that its value is invariant to changes in the scaling of M and Y. Precision in estimation of *MedES* can be evaluated by 95% CIs constructed via a bootstrapping procedure (Kraemer, 2014). These constitute a desirable set of properties for an effect size for the indirect effect for mediation models with binary predictors, and future research should further investigate the properties of *MedES*. For example, although SRDs are bounded by 0 and 1 (provided the treatment effect is in the expected direction), it is possible for differences in SRDs to be greater than 1 if the SRDs are of opposite sign, and less than 0 if the null SRD is larger than the overall SRD. It would be of interest to know the conditions under which *MedES* exceeds these boundaries and to determine the plausibility of these conditions in practice. It would also be of interest to investigate the sample properties of point and interval estimators of *MedES*. Overall, *MedES* appears to be a promising effect size measure for indirect effects for models with binary predictors, sharing several desirable properties with υ. Assuming the presently unknown properties of *MedES* are determined to be satisfactory, our view of which effect size is preferable would be largely informed by interpretability. However, as previously noted, the behavior of indirect effects in even the simplest mediation models can be complex and not readily captured by a single effect size measure, so we would recommend reporting both *MedES* and $\tilde{\upsilon}$ for these models.

Sample Size Planning

Notable among the reasons for using effect sizes outlined by the APA Task Force on Statistical Inference (Wilkinson & American Psychological Association Task Force on Statistical Inference, 1999) is sample size planning, so it is germane to discuss how υ may be used for this purpose. The most common approach to sample size planning in applied research is based on the NHST framework. Traditionally, sample size is determined using a computational formula that requires the specification of a hypothesized population effect size, the variance of the effect size estimator, the Type I error rate, and the desired level of power (typically .80). The estimator variance is derived from the sampling distribution, which is either known, as is the case for means (t-distributed) and variances (F-distributed), or is reasonably well approximated, as is the case for the indirect effect (Fritz & MacKinnon, 2007; Wang & Xue, 2016). The estimators of υ have neither known sampling distributions nor well-approximated sampling variances, so a computational formula for power analysis is not available.

An alternative approach to sample size planning in a power analytic framework is based on Monte Carlo simulation (Muthén & Muthén, 2002). This approach is particularly useful for power analysis in which the statistics of interest have unknown or intractable sampling distributions but are functions of statistics with known properties. Simulation-based power analysis works by fully specifying the parameters of the hypothesized population model, simulating a large number of data sets of a given sample size from the model, estimating the model parameters and computing statistics that are functions of those estimates, recording the statistical significance estimates in each data set, and computing power as the proportion of estimates significantly different from the respective null hypothesized values. Systematically increasing the sample size of the simulated data sets allows researchers to generate empirical power curves for any model parameter and determine the sample size necessary for the desired level of power. This approach can be implemented in common statistical software packages such as M*plus*, R, and SAS.

It is important to recall, however, that υ is not well-suited for NHSTs in which the null is zero because the null hypothesis is on the boundary of the parameter space. CIs for $\hat{\upsilon}$ are strictly positive and, although CIs for $\tilde{\upsilon}$ can contain values less than zero, negative estimates are inconsistent with the definition of the population parameter. Therefore, we would not recommend conducting power analysis using υ when the null hypothesis is of no effect. However, if the null hypothesis were some nonzero value, then simulation-based power analysis would be a promising approach for sample size planning using υ and its estimators. Because Monte Carlo simulation is a model-based method, it would be necessary to fully specify the parameters of the mediation model such that the model-derived υ has the desired effect size (e.g., small, medium, or large based on Cohen's standards for R^2). An additional benefit of this model-based approach to power analysis is that the relative magnitudes of the β_{MX} and $\beta_{YM \cdot X}$ coefficients can have differential influences on the CI width of the indirect effect and the sampling variances of the coefficients, which, as previously shown, influence $\tilde{\upsilon}$ estimates. However, it should be noted that a firm understanding of the phenomena under study is required to conduct power analyses with nonzero null hypotheses, and are uncommon in psychological research.

Although power analysis is standard practice in many applied research settings, there are alternative approaches to sample size planning for which υ may also be a useful effect size measure. One promising approach is the accuracy in parameter estimation (AIPE) framework (K. Kelley, 2007c, 2008; K. Kelley & Maxwell, 2003; K. Kelley & Rausch, 2006). Whereas power analysis fundamentally relies on the NHST framework, the focus of the AIPE approach to is to plan sample size for a desired CI width for the parameter of interest. There is a strong relationship between NHSTs and CIs, but the sample sizes required for the desired level of power compared with the desired CI width may differ substantially (K. Kelley, Maxwell, & Rausch, 2003; K. Kelley & Rausch, 2006), as they address fundamentally different questions. In essence, the goal of conducting a power analysis is to find a CI that excludes zero, whereas the goal of AIPE is to find a CI that is sufficiently narrow (regardless of whether or not zero is in the interval). AIPE sample size planning can be particularly useful for studies in which investigators have only vague knowledge about the model parameters. Power can vary substantially across combinations of parameter values, so if the true effect size differs from that for which the sample size was determined, the power of a study can be substantially less than the desired level. Planning for parameter precision is less dependent on the true value of the population parameter (Maxwell, Kelley, & Rausch, 2008).

Like sample size planning using power analysis, determining sample size using υ in the AIPE approach would also be based on Monte Carlo simulation. However, rather than specifying a desired level of power, the researcher must determine a desired maximum CI width with sufficient assurance that the proportion of CI widths larger than the specified maximum width is minimal (e.g., .05, .01). For example, one might select an assurance of .80, implying that 80% of 95% CI widths will be sufficiently narrow. The researcher would specify the mediation model corresponding to the desired υ effect size, simulate data sets for a given sample size, estimate and construct 95% CIs for $\tilde{\upsilon}$ in each data set, recording the

CI width, and computing the proportion of CI widths that are sufficiently narrow. The minimum necessary sample size is then the sample size for which the proportion of CI widths less than or equal to the maximum specified width is the specified assurance.

One can also combine power and AIPE in the Monte Carlo framework by determining the sample size that has at least the desired degree of power (e.g., .80) for an effect size parameter appropriate for NHSTs, while achieving a sufficient level of accuracy (i.e., a high probability of a confidence interval no wider than .10 units) for estimating υ. For example, it is presumable that a researcher using an effect size measure has already identified an indirect effect of interest via a significance test or otherwise. In this circumstance, the magnitude and precision of the υ estimate would be of more interest than its statistical significance. Applying this logic to sample size planning, a reasonable approach would be to combine a test of the existence of the indirect effect (i.e., can the null hypothesis that $B_{MX}B_{YM \cdot X} = 0$ be rejected?) and sufficient precision of the effect size estimate (i.e., is the CI width for $\tilde{\upsilon}$ less than or equal to the specified maximum CI width?). In such a combined approach, both power and accuracy are achieved, which is arguably a best practice for designing studies.

Limitations

Although υ is a promising effect size measure for mediation analysis, there are several limitations of the current study that should be considered prior to employing the effect size or its estimators in practice. First, results of the Monte Carlo simulation highlighted several circumstances in which percentile CIs for $\hat{\upsilon}$ and $\tilde{\upsilon}$ appear to be unstable. Specifically, coverage tended to be above a satisfactory level at smaller N and for smaller effect sizes for both estimators. In other words, 95% CIs constructed using the percentile bootstrap were overly wide (i.e., conservative) in these conditions. In addition, proportions of misses to the left and right of the CI were consistently imbalanced for $\tilde{\upsilon}$, such that the population value tended to be greater than the upper CI limit. It is possible that CIs constructed using other procedures (e.g., bias-corrected bootstrap, Monte Carlo) may be superior to percentile CIs in these circumstances, but a thorough investigation of these alternative CIs was beyond the scope of the current study. Second, the indirect effects considered in this study were restricted to simple three-variable variable mediation models, so the methods and results presented are not generalizable to the more complex mediation models (i.e., multiple mediators, moderators, covariates), but, in principle, they can be extended. Third, υ and its sample estimators were developed under the assumption that the mediation model illustrated in Figure 1 is correctly specified, and, therefore, the causal effect would be unbiased. However, as previously described, there are several assumptions necessary to justify claims that an indirect effect is a causal effect (i.e., constant effect, sequential ignorability), and violation of these assumptions introduces bias into estimation of the indirect effect. It follows, therefore, that assumption violations would also introduce bias into the estimation of υ. Although the resulting biases for indirect effects have been derived for many mediation models, the biases for υ are presently unknown. However, this does not prevent researchers from obtaining υ estimates from the bias-adjusted indirect effect estimates in a sensitivity analysis. It is assumed that, for a given set of conditions in a sensitivity analysis, the model is correctly specified, so the assumptions employed in the derivation of υ and its estimators are assumed to hold under these conditions as well (e.g., errors are uncorrelated given specified relationships with a confounder). Regarding the constant effect assumption, future work to extend υ to moderated mediation models is necessary to allow for obtaining effect size estimates when this assumption is relaxed. To be clear, however, there is nothing we assume here that researchers are not already assuming when using mediation models and hypothesizing a causal process.

Future Research

Future research should extend υ to these complex mediation models, including models with multiple parallel and sequential mediators, covariates, latent variables, moderators, and multilevel and longitudinal data. In addition, mediator and outcome variables considered in this study were continuous, and there may be alternative representations of υ for models with binary and count outcomes (e.g., pseudo-R^2). Future research could also address deficiencies in CI coverage and implement methods to construct stable CIs across a range of parameter combinations and Ns. Finally, research should evaluate the properties of υ when the assumptions of mediation analysis are violated. In concluding, we believe that υ advances research on mediation models by deriving a theoretically meaningful and highly useful effect size. Furthermore, we believe that $\tilde{\upsilon}$ has been demonstrated to be a quality estimator of υ and thus offers an advancement not only to the methods literature but also to users of mediation models.

References

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

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

Andrews, D. W. K. (2000). Inconsistency of the bootstrap when a parameter is on the boundary of the parameter space. *Econometrica, 68,* 399–405. http://dx.doi.org/10.1111/1468-0262.00114

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

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. http://dx.doi.org/10.1037/0022-3514.51.6.1173

Beasley, T. M. (2014). Tests of mediation: Paradoxical decline in statistical power as a function of mediator collinearity. *Journal of Experimental Education, 82,* 283–306. http://dx.doi.org/10.1080/00220973.2013.813360

Berry, K. J., & Mielke, P. W., Jr. (2002). Least sum of Euclidean regression residuals: Estimation of effect size. *Psychological Reports, 91,* 955–962. http://dx.doi.org/10.2466/pr0.2002.91.3.955

Blalock, H. M. (1962). Four-variable causal models and partial correlations. *American Journal of Sociology, 68,* 182–194. http://dx.doi.org/10.1086/223308

Bollen, K. A., & Stine, R. (1990). Direct and indirect effects: Classical and bootstrap estimates of variability. *Sociological Methodology, 20,* 115–140. http://dx.doi.org/10.2307/271084

Boomsma, A. (2013). Reporting Monte Carlo studies in structural equation modeling. *Structural Equation Modeling, 20,* 518–540. http://dx.doi.org/10.1080/10705511.2013.797839

Botha, J. D., Shapiro, A., & Steiger, J. H. (1988). Uniform indices-of-fit for factor analysis models. *Multivariate Behavioral Research, 23,* 443–450. http://dx.doi.org/10.1207/s15327906mbr2304_2

Bradley, J. V. (1978). Robustness? *British Journal of Mathematical and Statistical Psychology, 31,* 144–152. http://dx.doi.org/10.1111/j.2044-8317.1978.tb00581.x

Casella, G., & Berger, R. L. (2002). *Statistical inference*. Pacific Grove, CA: Duxbury Press.

Chernick, M. R. (2008). *Bootstrap methods: A guide for practitioners and researchers*. Hoboken, NJ: Wiley.

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

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

Cohen, J. (1994). The earth is round ($p < .05$). *American Psychologist, 49,* 997–1003. http://dx.doi.org/10.1037/0003-066X.49.12.997

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences*. New York, NY: Routledge.

de Heus, P. (2012). *R* squared effect-size measures and overlap between direct and indirect effect in mediation analysis. *Behavior Research Methods, 44,* 213–221. http://dx.doi.org/10.3758/s13428-011-0141-5

Dillon, W. R., & Goldstein, M. (1984). *Multivariate analysis*. New York, NY: Wiley.

Duncan, O. D. (1970). Partials, partitions, and paths. *Sociological Methodology, 2,* 38–47. http://dx.doi.org/10.2307/270782

Ezekiel, M. (1930). *Methods of correlational 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. http://dx.doi.org/10.3758/BRM.41.2.486

Fritz, M. S., & MacKinnon, D. P. (2007). Required sample size to detect the mediated effect. *Psychological Science, 18,* 233–239. http://dx.doi.org/10.1111/j.1467-9280.2007.01882.x

Fritz, M. S., Taylor, A. B., & Mackinnon, D. P. (2012). Explanation of two anomalous results in statistical mediation analysis. *Multivariate Behavioral Research, 47,* 61–87. http://dx.doi.org/10.1080/00273171.2012.640596

Goldin, P. R., Ziv, M., Jazaieri, H., Werner, K., Kraemer, H., Heimberg, R. G., & Gross, J. J. (2012). Cognitive reappraisal self-efficacy mediates the effects of individual cognitive-behavioral therapy for social anxiety disorder. *Journal of Consulting and Clinical Psychology, 80,* 1034–1040. http://dx.doi.org/10.1037/a0028555

Greenland, S., & Poole, C. (2013). Living with *p* values: Resurrecting a Bayesian perspective on frequentist statistics. *Epidemiology (Cambridge, Mass.), 24,* 62–68. http://dx.doi.org/10.1097/EDE.0b013e3182785741

Hoyle, R. H., & Kenny, D. A. (1999). Sample size, reliability, and tests of statistical mediation. In R. H. Hoyle (Ed.), *Statistical strategies for small sample research* (pp. 195–222). Thousand Oaks, CA: Sage.

Imai, K., Keele, L., & Tingley, D. (2010). A general approach to causal mediation analysis. *Psychological Methods, 15,* 309–334. http://dx.doi.org/10.1037/a0020761

Imai, K., Keele, L., & Yamamoto, T. (2010). Identification, inference and sensitivity analysis for causal mediation effects. *Statistical Science, 25,* 51–71. http://dx.doi.org/10.1214/10-STS321

Jöreskog, K. G. (1999). *How large can a standardized coefficient be?* Retrieved from http://www.ssicentral.com/lisrel/column2.htm

Kelley, K. (2007a). Confidence intervals for standardized effect sizes: Theory, application, and implementation. *Journal of Statistical Software, 20,* 1–24. http://dx.doi.org/10.18637/jss.v020.i08

Kelley, K. (2007b). Methods for the behavioral, educational, and social sciences: An R package. *Behavior Research Methods, 39,* 979–984. http://dx.doi.org/10.3758/BF03192993

Kelley, K. (2007c). Sample size planning for the coefficient of variation from the accuracy in parameter estimation approach. *Behavior Research Methods, 39,* 755–766. http://dx.doi.org/10.3758/BF03192966

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. http://dx.doi.org/10.1080/00273170802490632

Kelley, K. (2017). MBESS (Version 4.4.0 or later) [Computer software and manual]. Retrieved from https://CRAN.R-project.org/package=MBESS

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. http://dx.doi.org/10.1037/1082-989X.8.3.305

Kelley, K., Maxwell, S. E., & Rausch, J. R. (2003). Obtaining power or obtaining precision. Delineating methods of sample-size planning. *Evaluation & the Health Professions, 26,* 258–287. http://dx.doi.org/10.1177/0163278703255242

Kelley, K., & Preacher, K. J. (2012). On effect size. *Psychological Methods, 17,* 137–152. http://dx.doi.org/10.1037/a0028086

Kelley, K., & Rausch, J. R. (2006). Sample size planning for the standardized mean difference: Accuracy in parameter estimation via narrow confidence intervals. *Psychological Methods, 11,* 363–385. http://dx.doi.org/10.1037/1082-989X.11.4.363

Kelley, T. L. (1935). An unbiased correlation ratio measure. *Proceedings of the National Academy of Sciences, 21,* 554–559. http://dx.doi.org/10.1073/pnas.21.9.554

Keuzenkamp, H. A., & Magnus, J. R. (1995). On tests and significance in econometrics. *Journal of Econometrics, 67,* 5–24. http://dx.doi.org/10.1016/0304-4076(94)01624-9

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

Kraemer, H. C. (2014). A mediator effect size in randomized clinical trials. *International Journal of Methods in Psychiatric Research, 23,* 401–410. http://dx.doi.org/10.1002/mpr.1445

Kraemer, H. C., & Kupfer, D. J. (2006). Size of treatment effects and their importance to clinical research and practice. *Biological Psychiatry, 59,* 990–996. http://dx.doi.org/10.1016/j.biopsych.2005.09.014

MacKinnon, D. P. (2008). *Introduction to statistical mediation analysis*. New York, NY: Taylor & Francis.

MacKinnon, D. P., Fritz, M. S., Williams, J., & Lockwood, C. M. (2007). Distribution of the product confidence limits for the indirect effect: Program PRODCLIN. *Behavior Research Methods, 39,* 384–389. http://dx.doi.org/10.3758/BF03193007

MacKinnon, D. P., Krull, J. L., & Lockwood, C. M. (2000). Equivalence of the mediation, confounding and suppression effect. *Prevention Science, 1,* 173–181. http://dx.doi.org/10.1023/A:1026595011371

MacKinnon, D. P., Lockwood, C. M., & Williams, J. (2004). Confidence limits for the indirect effect: Distribution of the product and resampling methods. *Multivariate Behavioral Research, 39,* 99–128. http://dx.doi.org/10.1207/s15327906mbr3901_4

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

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. http://dx.doi.org/10.1146/annurev.psych.59.103006.093735

Morrison, D. F., & Henkel, R. E. (1970). *The significance test controversy—A reader*. Chicago, IL: Aldine.

Muthén, L. K., & Muthén, B. O. (2002). How to use a Monte Carlo study to decide on sample size and determine power. *Structural Equation Modeling, 9,* 599–620. http://dx.doi.org/10.1207/S15328007SEM0904_8

Olkin, E., & Pratt, J. W. (1958). Unbiased estimation of certain correlation coefficients. *Annals of Mathematical Statistics, 29,* 201–211. http://dx.doi.org/10.1214/aoms/1177706717

Pearl, J. (2014). Interpretation and identification of causal mediation. *Psychological Methods, 19,* 459–481. http://dx.doi.org/10.1037/a0036434

Preacher, K. J., & Hayes, A. F. (2008). 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. http://dx.doi.org/10.4135/9781452272054.n2

Preacher, K. J., & Kelley, K. (2011). Effect size measures for mediation models: Quantitative strategies for communicating indirect effects. *Psychological Methods, 16,* 93–115. http://dx.doi.org/10.1037/a0022658

R Development Core Team. (2014). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from http://www.R-project.org/

Rucker, D. D., Preacher, K. J., Tormala, Z. L., & Petty, R. E. (2011). Mediation analysis in social psychology: Current practices and new recommendations. *Social and Personality Psychology Compass, 5,* 359–371. http://dx.doi.org/10.1111/j.1751-9004.2011.00355.x

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

Simon, H. (1954). Spurious correlation: A causal interpretation. *Journal of the American Statistical Association, 49,* 467–479.

Sobel, E. (1982). Asymptotic confidence intervals for indirect effects in structural equation models. *Sociological Methodology, 13,* 290–312. http://dx.doi.org/10.2307/270723

Tofighi, D., & MacKinnon, D. P. (2011). RMediation: An R package for mediation analysis confidence intervals. *Behavior Research Methods, 43,* 692–700. http://dx.doi.org/10.3758/s13428-011-0076-x

Vanderweele, T. J. (2015). *Explanation in causal inference: Methods for mediation and interaction*. New York, NY: Oxford University Press.

Venables, W. N., & Ripley, B. D. (2002). *Modern applied statistics with S* (4th ed.). New York, NY: Springer. http://dx.doi.org/10.1007/978-0-387-21706-2

Wang, C., & Xue, X. (2016). Power and sample size calculations for evaluating mediation effects in longitudinal studies. *Statistical Methods in Medical Research, 25,* 686–705. http://dx.doi.org/10.1177/0962280212465163

Wen, Z., & Fan, X. (2015). Monotonicity of effect sizes: Questioning kappa-squared as mediation effect size measure. *Psychological Methods, 20,* 193–203. http://dx.doi.org/10.1037/met0000029

Wilkinson, L., & American Psychological Association Task Force on Statistical Inference. (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist, 54,* 594–604. http://dx.doi.org/10.1037/0003-066X.54.8.594

Wright, S. (1960). Path coefficients and path regressions: Alternative or complementary concepts? *Biometrics, 16,* 189–202. http://dx.doi.org/10.2307/2527551

Yin, P., & Fan, X. (2001). Estimating R^2 shrinkage in multiple regression: A comparison of different analytic methods. *Journal of Experimental Education, 69,* 203–224. http://dx.doi.org/10.1080/00220970109600656

Yuan, Y., & MacKinnon, D. P. (2009). Bayesian mediation analysis. *Psychological Methods, 14,* 301–322. http://dx.doi.org/10.1037/a0016972

Received November 12, 2016
Revision received August 12, 2017
Accepted August 25, 2017 ■