Psychological Methods
2006, Vol. 11, No. 2, 142–163

1082-989X/06/$12.00 DOI: 10.1037/1082-989X.11.2.142

Conceptualizing and Testing Random Indirect Effects and Moderated Mediation in Multilevel Models: New Procedures and Recommendations

Daniel J. Bauer, Kristopher J. Preacher, and Karen M. Gil
University of North Carolina at Chapel Hill

The authors propose new procedures for evaluating direct, indirect, and total effects in multilevel models when all relevant variables are measured at Level 1 and all effects are random. Formulas are provided for the mean and variance of the indirect and total effects and for the sampling variances of the average indirect and total effects. Simulations show that the estimates are unbiased under most conditions. Confidence intervals based on a normal approximation or a simulated sampling distribution perform well when the random effects are normally distributed but less so when they are nonnormally distributed. These methods are further developed to address hypotheses of moderated mediation in the multilevel context. An example demonstrates the feasibility and usefulness of the proposed methods.

Keywords: multilevel model, hierarchical linear model, indirect effect, mediation, moderated mediation

In psychology and other social sciences, hypotheses often concern the causal pathways through which key predictors transmit their effects to specific outcomes. For example, Wei, Mallinckrodt, Russell, and Abraham (2004) examined the extent to which attachment anxiety and avoidance were related to maladaptive perfectionism and, in turn, depressive mood. Similarly, Catanzaro and Laurent (2004) hypothesized that alcohol expectancies play a role in drinking as a means of coping, which then predicts a number of drinking behaviors. In both of these cases, the causal effects of the original predictor are transmitted at least partially through an intervening variable, as diagrammed in Figure 1. In the diagram, the product of the paths labeled a and b represents the *indirect effect* of X on Y, the path labeled c' represents the *direct effect* of X on Y, and $c = ab + c'$ is the *total effect* of X on Y (Alwin & Hauser, 1975; Bollen, 1987, 1989). Depending on the pattern of these effects, the variable M may be called a mediator, a suppressor, or simply an intervening variable (MacKinnon, Krull, & Lockwood, 2000). We generally use the terms *mediator* and *mediation* to be consistent with the focus of much of the literature on indirect effects, although our exposition also pertains to other patterns of indirect effects.

Within the context of linear regression and path analysis, a number of methods have been proposed for evaluating mediation, and this remains an active area of research (e.g., MacKinnon, Lockwood, Hoffman, West, & Sheets, 2002; MacKinnon, Lockwood, & Williams, 2004; Shrout & Bolger, 2002). Such methods are, however, inappropriate if the analyzed data are actually hierarchical in nature (i.e., composed of two or more nested levels). Two forms of hierarchical data are common in psychological research. First, individuals may be assessed from a number of groups. Individuals are then said to be nested within groups. To the extent that individuals within a group share common experiences, we would expect their scores on the outcome variable to be correlated across members of the group, which violates the independence assumption of many statistical models. Second, repeated observations may be made on the same individuals. In this case, the repeated measurements are said to be nested within individuals. Repeated measurements are typically correlated within persons, which again

Daniel J. Bauer, Kristopher J. Preacher, and Karen M. Gil, Department of Psychology, University of North Carolina at Chapel Hill.

Additional materials are on the Web at http://dx.doi.org/10.1037/1082-989X.11.2.142.supp

This work was funded in part by National Institute on Drug Abuse Grant F32 DA16883 awarded to Kristopher J. Preacher. In addition, the study of sickle cell disease that formed the basis of the applied analyses was supported by National Institutes of Health Grant R01 HL6217 awarded to Karen M. Gil, principal investigator. We also thank Greg Stonerock for assisting in the preparation of the sickle cell disease data and Li Cai for assisting with the technical appendix.

Correspondence concerning this article should be addressed to Daniel J. Bauer, Department of Psychology, Davie Hall 337-A, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599-3270. E-mail: dbauer@email.unc.edu

compromises the independence assumption. For both types of hierarchical data, we refer to the lower level as Level 1 and the upper level as Level 2. More than two levels of data are possible, but we restrict our attention to two-level models.

Because the independence assumption is violated for these data structures, multiple linear regression and path analysis will produce biased tests of the effects in the model (Hox, 2002; Kreft & de Leeuw, 1998; Raudenbush & Bryk, 2002). One way to appropriately model such data is to use a multilevel model, also known as a hierarchical linear model or a mixed-effects model. A key advantage of the multilevel model is that it captures the correlations among the Level 1 observations through the estimation of random effects. These random effects can take the form of random intercepts, reflecting differences in the overall level of the outcome variable across Level 2 units; random slopes, reflecting differences in the effects of predictors across Level 2 units; or both. For instance, a multilevel model with a random intercept produces a compound symmetric correlation structure for the Level 1 observations. In addition to the statistical advantages of multilevel models for correlated data, random effects can have quite interesting substantive interpretations. A random slope, for example, indicates that the causal effect of the predictor on the outcome differs over Level 2 units. This may then initiate the search for potential moderators of the effect.

Another important feature of hierarchical data is that predictors can reside at different levels of the data (e.g., individual- vs. group-level characteristics). Given this, mediation in multilevel models may take several forms, as shown in Figure 2 (Kenny, Kashy, & Bolger, 1998; Krull & MacKinnon, 2001). *Upper level mediation* exists when the effect of a Level 2 predictor on a Level 1 outcome is mediated by another Level 2 predictor (2 → 2 → 1 mediation). *Lower level mediation* exists when the mediator is a Level 1 variable. In some cases of lower level mediation the effect of a Level 2 predictor is mediated (2 → 1 → 1 mediation), and in other cases the effect of a lower level predictor is mediated (1 → 1 → 1 mediation).[1] Methods for assessing upper level mediation have been proposed by

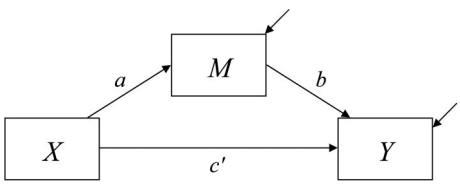

Figure 1. Diagram of a simple mediation model. Variables are indicated in boxes: *X* is the primary predictor, *M* is the mediator, and *Y* is the distal outcome variable. Arrows originating from variables indicate hypothesized causal effects. Labels for these effects are indicated next to the arrows. Arrows not originating from variables indicate residuals.

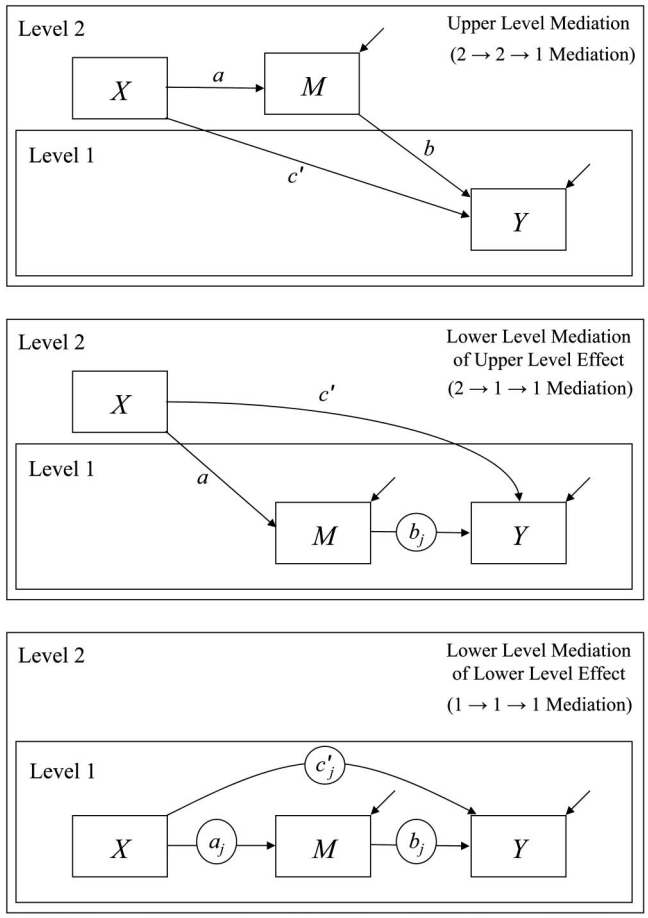

Figure 2. Upper and lower level mediation in a two-level model: Nested frames indicate levels of sampling, boxes indicate variables, arrows without circles represent fixed effects, arrows with circles represent random effects, and arrows not originating from variables indicate residuals.

Raudenbush and Sampson (1999) and Bauer (2003). Krull and MacKinnon (1999, 2001) offered an alternative method applicable to all three types of multilevel mediation that is similar to the causal steps approach of Baron and Kenny (1986) for ordinary regression models. More recently, Pituch, Wittaker, and Stapleton (2005) examined several approaches for testing indirect effects in 2 → 2 → 1 and 2 → 1 → 1 models for cluster-randomized treatment designs.

What all of these approaches hold in common is an assumption that the causal effects are fixed (not random),

[1] As a reviewer pointed out, one could also consider other forms of mediation for multilevel data. For instance, one might hypothesize that an individual-level predictor affects a group-level mediator, which, in turn, affects a group- or individual-level outcome. In this article, we consider only models in which the causal effects reside at the same level or are transmitted from a higher to a lower level variable.

meaning that the magnitude of the effects is equal for all Level 2 units. In lower level mediation, however, the causal effects can be random because some predictors reside at Level 1. In $2 \rightarrow 1 \rightarrow 1$ mediation, the effect of the mediator on the outcome may be random; in $1 \rightarrow 1 \rightarrow 1$ mediation, all three causal effects can be random. These random effects are indicated in Figure 2 as circles on the causal paths. They represent heterogeneity in the causal effects across Level 2 units and may be of considerable substantive interest. For instance, in the empirical analysis to be presented later, we analyze daily diary data to assess whether emotional reactions to pain mediate the relation between physical pain and stress in patients with sickle cell disease (SCD) and whether the strength of mediation differs across participants.

Kenny, Korchmaros, and Bolger (2003) first called attention to the importance of random effects in lower level mediation models, in particular in $1 \rightarrow 1 \rightarrow 1$ models. Furthermore, they proposed a method for fitting $1 \rightarrow 1 \rightarrow 1$ models when all three causal pathways are random effects, as depicted in the lower panel of Figure 2. The purpose of the present article is to expand on this prior research to improve the conceptualization and estimation of lower level mediation with random effects in multilevel models. We begin by showing how one can estimate the $1 \rightarrow 1 \rightarrow 1$ model simultaneously using conventional multilevel or mixed modeling software. We then clarify that, in the presence of random indirect effects, one may be interested in two related but different questions. First, is there heterogeneity in the strength of the indirect, direct, and total effects across units of the population? Second, how precisely can we estimate the average effects in the population? After distinguishing these questions, we proceed to a small simulation study designed to evaluate the quality of the average effect estimates and their confidence intervals (CIs). Next, we extend the model to include predictors of heterogeneity in the indirect and direct effects. Last, we demonstrate the value of estimating and predicting random indirect effects in $1 \rightarrow 1 \rightarrow 1$ models with an empirical example. Our focus throughout is on the estimation and testing of the indirect, direct, and total effects of the distal predictor on the outcome, not on the broader causal steps approach to mediation outlined by Baron and Kenny (1986).

Lower Level Mediation Models With Random Indirect Effects

Loosely following the notation of Kenny et al. (2003), one can write the lower level mediation model depicted in the lower panel of Figure 2 with two Level 1 equations as

$$M_{ij} = d_{Mj} + a_j X_{ij} + e_{Mij}$$

$$Y_{ij} = d_{Yj} + b_j M_{ij} + c'_j X_{ij} + e_{Yij}. \quad (1)$$

The terms e_{Mij} and e_{Yij} are Level 1 residuals for M and Y, respectively. The other five terms are interpreted similarly to the intercepts and slopes of a standard regression model, with the caveat that each coefficient is random, meaning that the value of the coefficient varies across Level 2 units (as indicated by the j subscript). That is, the intercepts for M and Y are designated d_{Mj} and d_{Yj}, respectively, the effect of X on M is designated a_j, the effect of M on Y is designated b_j, and the direct effect of X on Y is designated c'_j. The random effects of the model permit heterogeneity in the causal effects. For instance, if Level 1 represents repeated measures, indexed by i, and Level 2 represents persons, indexed by j, then this model can be used to assess how the strength of the hypothesized causal relations among X, M, and Y vary across individuals. For simplicity, we assume for the time being that there are no Level 2 predictors of these random effects.

A number of assumptions are required to make the model estimable and ensure that the effects are unbiased. These assumptions are inclusive of those that apply to all multilevel models (as conventionally estimated) but include features that are unique to multivariate multilevel models, such as the lower level mediation model in Equation 1. These unique features are italicized in the following list of assumptions:[2]

1. The predictors are uncorrelated with the random effects (intercepts and slopes) and residuals, both within *and across equations* (e.g., X_{ij} must be uncorrelated with d_{Mj}, a_j, and e_{Mj} and uncorrelated with d_{Yj}, b_j, c'_j, and e_{Yij}).

2. The residuals e_{Mj} and e_{Yij} are each normally distributed with an expected value of zero, *and they are uncorrelated with one another*. Typically, the residuals for each outcome are assumed to be independent and homoscedastic across i within j, but these restrictions can and should be relaxed in certain circumstances (e.g., when residuals are expected to autocorrelate with repeated measures). The additional assumption that the residuals are uncorrelated across outcomes is required to identify the effect of M on Y.

3. The random effects are normally distributed with means equal to the average effects in the population. Although other assumptions are possible, we also assume that

[2] Another assumption that is sometimes listed for these models is that the predictors are fixed. This assumption seems to present a problem for the model in Equation 1 because M appears in the second equation as a fixed predictor but also appears in the first equation as a random outcome variable. Like ordinary regression, however, the assumption that the predictors are fixed is unnecessary as long as the predictors are uncorrelated with the errors, as we note in Assumption 1 (Demidenko, 2004, p. 143). We can thus consider M to be a random variable in both equations. Consequently, we must now assume a distribution for M, namely that it is conditionally normal (conditional on the random effects and X), as indicated in Assumption 5.

all of the random effects covary with one another. (Note that the random effects of a multilevel model are often expressed in deviation form, with the means subtracted, so that the expected values of the random effects are zero.)

4. The Level 1 residuals are uncorrelated with the random effects both within *and across equations* (e.g., e_{Mj} is uncorrelated with d_{Mj}, a_j, d_{Yj}, b_j, and c_j').

5. Assumptions 2 and 3 imply that the distributions of M and Y are conditionally normal. That is, the distribution of M is normal conditional on X and the distribution of Y is normal conditional on M and X.[3]

We show that Assumption 3 is particularly important for deriving estimates of the indirect effects. We can write Assumption 3 more formally as

$$\begin{pmatrix} d_{Mj} \\ a_j \\ d_{Yj} \\ b_j \\ c_j' \end{pmatrix} \sim N\left[\begin{pmatrix} d_M \\ a \\ d_Y \\ b \\ c' \end{pmatrix}, \begin{pmatrix} \sigma^2_{d_{Mj}} & & & & & \\ \sigma_{d_{Mj},a_j} & \sigma^2_{a_j} & & & & \\ \sigma_{d_{Mj},d_{Yj}} & \sigma_{a_j,d_{Yj}} & \sigma^2_{d_{Yj}} & & & \\ \sigma_{d_{Mj},b_j} & \sigma_{a_j,b_j} & \sigma_{d_{Yj},b_j} & \sigma^2_{b_j} & & \\ \sigma_{d_{Mj},c_j'} & \sigma_{a_j,c_j'} & \sigma_{d_{Yj},c_j'} & \sigma_{b_j,c_j'} & \sigma^2_{c_j'} & \end{pmatrix}\right]. \quad (2)$$

The means of the random effects are designated with the same symbols as the random effects, absent the subscript j to indicate they are constant over Level 2 units (e.g., a is the mean of a_j). These are the fixed effects of the model. The covariance matrix for the random effects captures the nonindependence of the observations collected at Level 1 and, together with the variances of the Level 1 residuals, constitutes the variance components of the model. The diagonal elements of this matrix, the variances, characterize heterogeneity in the causal effects of the model across Level 2 units. For instance, $\sigma^2_{c_j'}$ is the variance of the direct effect (c_j'), indicating the extent to which the strength of the direct effect differs over Level 2 units. The off-diagonal elements of the matrix capture the covariance among the random effects. For instance, a positive value for σ_{a_j,b_j} would indicate that for Level 2 units for which the effect of X on M (a_j) is high, the effect of M on Y (b_j) is often also high.

To examine lower level mediation, Kenny et al. (2003) estimated each Level 1 model in Equation 1 separately. Unfortunately, this two-step approach sacrifices some important information. First, one cannot estimate directly the covariance of random effects residing in different Level 1 models, including all of the elements enclosed in the dotted square in Equation 2. As we show, this makes it difficult to capture the dispersion of the indirect and total effects of the model. Kenny et al. (2003) proposed an ad hoc solution to the problem that involved estimating ordinary least squares regression models for M and Y separately for each unit j, collecting the slope estimates, and then calculating the sample covariance of the slope estimates. They acknowledged, however, that this was not an optimal strategy. A second limitation of the two-step approach is that the asymptotic covariance matrix of the model estimates is similarly incomplete. This matrix captures the sampling variability in the estimates, and some of its missing elements are needed to capture the standard errors of the average indirect and total effects in the model.

The solution that we propose is to formulate the model with a single Level 1 equation through the use of selection (or indicator) variables. Researchers have used this strategy in the past to aid in the estimation of other kinds of multivariate multilevel models (e.g., MacCallum, Kim, Malarkey, & Kiecolt-Glaser, 1997), and it is equally useful in this context. The basic idea is to form a new outcome variable—for instance, Z—by stacking Y and M for each unit i within j. This single outcome variable allows us to fit a "multivariate" model using univariate multilevel modeling software. To distinguish the two variables stacked in Z, we also create two selection variables—for instance, S_M and S_Y. The variable S_M is set equal to 1 when Z refers to M and is 0 otherwise. Similarly, the variable S_Y is set equal to 1 when Z refers to Y and is 0 otherwise. We retain the variables X and M in the new data set, as they are needed as predictors of Z. An example of this data rearrangement is presented in Figure 3.

The purpose of rearranging the data in this fashion is that we can now specify the lower level mediation model with a single equation:

$$Z_{ij} = S_{M_{ij}}(d_{Mj} + a_j X_{ij}) + S_{Y_{ij}}(d_{Yj} + b_j M_{ij} + c_j' X_{ij}) + e_{Zij}. \quad (3)$$

Notice that the two selection variables essentially toggle from the model for M to the model for Y in Equation 3. For instance, when Z is a value of M, then $S_M = 1$ and $S_Y = 0$, and Equation 3 simplifies to $Z_{ij} = d_{Mj} + a_j X_{ij} + e_{Zij}$. This is the model for M from Equation 1. Similarly, when Z is a value of Y, then $S_Y = 1$ and $S_M = 0$, and Equation 3 simplifies to $Z_{ij} = d_{Yj} + b_j M_{ij} + c_j' X_{ij} + e_{Zij}$, or the model for Y from Equation 1. Thus, although we have stacked Y and M into a single dependent variable for the purpose of

[3] A reviewer correctly noted that the unconditional joint distribution of M and Y would not be normal, given the contribution of the $a_j b_j$ product to Y. However, it is important to note that the conditional distribution of M given X and Y given M and X can still be normal and that it is this conditional distribution for which normality is assumed. Although this assumption is common to all linear multilevel models, it may often be unrealistic in practice. Therefore, in our simulation study, we consider the consequences of violating this assumption when the random effects are nonnormally distributed.

j	Y	M	X
1	0.57	0.11	1.55
1	1.21	2.11	2.28
⋮	⋮	⋮	⋮
1	0.23	0.73	0.92
2	-1.15	-0.36	1.00
2	-3.72	-2.97	-1.19
⋮	⋮	⋮	⋮
2	-3.86	-2.56	-1.56

Restructuring the data

j	Z	S_Y	S_M	M	X
1	0.57	1	0	0.11	1.55
1	0.11	0	1	0.11	1.55
1	1.21	1	0	2.11	2.28
1	2.11	0	1	2.11	2.28
⋮	⋮	⋮	⋮	⋮	⋮
1	0.23	1	0	0.73	0.92
1	0.73	0	1	0.73	0.92
2	-1.15	1	0	-0.36	1.00
2	-0.36	0	1	-0.36	1.00
2	-3.72	1	0	-2.97	-1.19
2	-2.97	0	1	-2.97	-1.19
⋮	⋮	⋮	⋮	⋮	⋮
2	-3.86	1	0	-2.56	-1.56
2	-2.56	0	1	-2.56	-1.56

Figure 3. In the top diagram, data are shown for two upper level units ($j = 1$ and $j = 2$) for an outcome (Y), a mediator (M), and a predictor (X) whose effect on Y is thought to be mediated by M. This data set is then rearranged into the lower diagram through the creation of a new variable Z that represents Y whenever the selection variable $S_Y = 1$ (and $S_M = 0$) and represents M whenever the selection variable $S_M = 1$ (and $S_Y = 0$). For each observation in the top diagram, there are then two rows in the bottom diagram. The values of M and X are repeated for these two rows.

fitting the model, the two outcomes continue to be distinguished in the model equations by the selection variables.

To aid in specifying the model, it is helpful to distribute the selection variables in Equation 3 as follows:

$$Z_{ij} = d_{Mj}S_{Mij} + a_j(S_{Mij}X_{ij}) + d_{Yj}S_{Y_{ij}} + b_j(S_{Y_{ij}}M_{ij}) + c'_j(S_{Y_{ij}}X_{ij}) + e_{Zij}. \quad (4)$$

Equation 4 shows that one would specify a model for Z with no intercept but with random effects for S_M and S_Y (d_{Mj} and d_{Yj}, respectively) and with random effects for the product variables S_MX, S_YM, and S_YX (a_j, b_j, and c'_j, respectively). In addition, one must use some method to allow the residual variance Var(e_{Zij}) to differ depending on S_M (or, equivalently, S_Y). This represents a form of heteroscedasticity because the residual variance for Z is then conditional on S_M. Fortunately, most multilevel modeling software programs offer one or more options for modeling heteroscedasticity. Depending on the substantive context, more complicated Level 1 residual variance (or covariance) structures can also be entertained (see the example we provide later in the article). Generic syntax for rearranging the data and fitting the model in Equation 4 in SAS is provided on the Web at http://dx.doi.org/10.1037/1082-989X.11.2.142.supp, although other programs, such as SPSS, HLM, or MLwiN, could also be used to fit the model.

Note that this specification strategy generalizes to models with more than one mediator through the creation of additional selection variables. The key is to "trick" the software into estimating a multivariate system of equations through the creation of a single outcome variable Z. Given this set-up, one can estimate the entire model simultaneously using conventional software for univariate multilevel models. The resulting output includes the full covariance matrix of the random effects and the full asymptotic covariance matrix of the fixed effects estimates and variance components, both of which are critical for our next developments.

To summarize to this point, we note that by using the selection variable approach, we can estimate the complete lower level mediation model simultaneously, providing all of the necessary information for evaluating the hypothesized causal effects of the model. We now turn to methods for quantifying and testing variability in these causal effects.

Evaluating Random Indirect and Total Effects in Multilevel Models

In this section, we consider two related questions: Given the presence of random effects in a lower level mediation model, how does one quantify variability in the indirect and total effects across Level 2 units of the population, and how does one obtain CIs for the average indirect and total effects? Although these are related questions, they are quite different in purpose. For the first question, we are interested in variability across Level 2 units (i.e., heterogeneity in causal effects). For the second question, we are interested in variability in our estimates of the average effects across Level 2 units (i.e., the precision of the average causal effects). This distinction is analogous to the difference between the standard deviation of a variable and the standard error of its mean. We now present the technical details for evaluating each question. Additional details on the derivation of the results presented in this section are provided in the Appendix. Readers who wish to see only the conclusions of these developments can skip to the *Summary of Procedures* subsection.

Investigating Heterogeneity in Causal Effects

The first question concerns how to characterize heterogeneity in the causal effects over Level 2 units. Let us consider the indirect effect first. The indirect effect for a given unit j is a_jb_j. As Kenny et al. (2003) noted, because a_j and b_j are not necessarily independent, the expected value or average of a_jb_j is (Goodman, 1960, p. 712):

$$E(a_j b_j) = ab + \sigma_{a_j,b_j}. \quad (5)$$

As such, the average indirect effect in the population is a function of the average effect of X on M (or a), the average effect of M on Y (or b), and the covariance between the two random effects (or σ_{a_j,b_j}). One interesting implication of this formula is that even if a and b are zero, the average indirect effect can be nonzero through the contribution of σ_{a_j,b_j}. We note that this formula does not require any assumptions on the distributions of the random effects (Kendall & Stuart, 1969, p. 283).

Making the assumption that a_j and b_j are normally distributed, Kenny et al. (2003) showed that the variance of $a_j b_j$ is

$$\mathrm{Var}(a_j b_j) = b^2\sigma^2_{a_j} + a^2\sigma^2_{b_j} + \sigma^2_{a_j}\sigma^2_{b_j} + 2ab\sigma_{a_j,b_j} + \sigma^2_{a_j,b_j}. \quad (6)$$

This variance quantifies heterogeneity in the strength of the indirect effect at Level 2. This value may be of considerable interest, as it indicates the extent to which the indirect effect of X on Y varies across the Level 2 sampling units. Similarly, Kenny et al. (2003) noted that the average total effect, $a_j b_j + c'_j$, can be expressed as

$$E(a_j b_j + c'_j) = ab + \sigma_{a_j,b_j} + c'. \quad (7)$$

Again with the assumption of normality of the random effects, the variance of the random total effect can be expressed as

$$\mathrm{Var}(a_j b_j + c'_j) = b^2\sigma^2_{a_j} + a^2\sigma^2_{b_j} + \sigma^2_{a_j}\sigma^2_{b_j} + 2ab\sigma_{a_j,b_j} + \sigma^2_{a_j,b_j} + \sigma^2_{c'_j} + 2b\sigma_{a_j,c'_j} + 2a\sigma_{b_j,c'_j}. \quad (8)$$

This variance, in turn, quantifies heterogeneity at Level 2 in the strength of the total effect.

One nice property of maximum likelihood estimation, the typical method for fitting multilevel models, is that any function of maximum likelihood estimates (MLEs) is itself an MLE (Raudenbush & Bryk, 2002, p. 52). As such, we can obtain MLEs for the quantities in Equations 5 through 8 by inserting the estimates from the fitted model in place of their population values. The simultaneous modeling approach we have described generates all of these estimates from a single model. Of particular interest may be the estimated average indirect and total effects, which we can obtain by inserting the model estimates into Equations 5 and 7. Our second question is concerned with the precision of these estimated average indirect and total effects.

Quantifying the Precision of the Estimated Average Causal Effects

We begin by emphasizing that Equations 6 and 8 capture heterogeneity in the indirect effects and total effects in the population of Level 2 units.[4] If we are instead interested in the precision of the estimated average indirect effect, $\hat{a}\hat{b} + \hat{\sigma}_{a_j,b_j}$, and estimated average total effect, $\hat{a}\hat{b} + \hat{\sigma}_{a_j,b_j} + \hat{c}'_j$, then we can calculate the sampling variances of these estimates as

$$\mathrm{Var}(\hat{a}\hat{b} + \hat{\sigma}_{a_j,b_j}) = b^2\mathrm{Var}(\hat{a}) + a^2\mathrm{Var}(\hat{b}) + \mathrm{Var}(\hat{a})\mathrm{Var}(\hat{b}) + 2ab\mathrm{Cov}(\hat{a}, \hat{b}) + \mathrm{Cov}(\hat{a}, \hat{b})^2 + \mathrm{Var}(\hat{\sigma}_{a_j,b_j}) \quad (9)$$

and

$$\mathrm{Var}(\hat{a}\hat{b} + \hat{\sigma}_{a_j,b_j} + \hat{c}') = b^2\mathrm{Var}(\hat{a}) + a^2\mathrm{Var}(\hat{b}) + \mathrm{Var}(\hat{a})\mathrm{Var}(\hat{b}) + 2ab\mathrm{Cov}(\hat{a}, \hat{b}) + \mathrm{Cov}(\hat{a}, \hat{b})^2 + \mathrm{Var}(\hat{c}') + 2b\mathrm{Cov}(\hat{a}, \hat{c}') + 2a\mathrm{Cov}(\hat{b}, \hat{c}') + \mathrm{Var}(\hat{\sigma}_{a_j,b_j}). \quad (10)$$

The variances and covariances in these equations (designated Var and Cov) represent the asymptotic sampling variances and covariances of the fixed effect estimates $\hat{a}$, $\hat{b}$, and $\hat{c}'$ and the covariance estimate $\hat{\sigma}_{a_j,b_j}$.

In practice, one must replace the population values in Equations 9 and 10 with their sample estimates (the model estimates and estimated asymptotic variances and covariances of the estimates) to obtain the estimated sampling variances of the average direct and indirect effects. Most multilevel or mixed modeling software programs provide an option to output the estimated variance–covariance matrix of the model estimates. Note that only by estimating the full model simultaneously can one estimate covariances between fixed effects residing in different equations—for example, $\mathrm{Cov}(\hat{a}, \hat{b})$ and $\mathrm{Cov}(\hat{a}, \hat{c}')$.

To make inferences concerning the average indirect and total effects, we can form CIs for the estimates. One method for constructing CIs assumes normality for the sampling distributions of the estimates. Under this assumption, 95% CIs for the average indirect effect and average total effect are obtained as

$$(\hat{a}\hat{b} + \hat{\sigma}_{a_j,b_j}) \pm 1.96\,[\widehat{Var}(\hat{a}\hat{b} + \hat{\sigma}_{a_j,b_j})]^{1/2} \quad (11)$$

and

$$(\hat{a}\hat{b} + \hat{\sigma}_{a_j,b_j} + \hat{c}') \pm 1.96\,[\widehat{Var}(\hat{a}\hat{b} + \hat{\sigma}_{a_j,b_j} + \hat{c}')]^{1/2}, \quad (12)$$

where ± 1.96 is the critical value of the z distribution and $\widehat{\mathrm{Var}}$ is used to indicate the estimated sampling variance obtained when the sample-based estimates are inserted into

[4] Kenny et al. (2003) sometimes interpreted these equations to give the sampling variances of the estimated average indirect and total effects.

Equation 9 or 10. Alternatively, one could perform a null hypothesis test by forming the critical ratio of each estimate to its standard error and comparing the result with the critical value of the z distribution. In either case, the assumption of normality for the sampling distributions will not hold exactly, given that $\hat{a}\hat{b}$ is a product of normally distributed estimates (and hence will not be normal). The deviation from normality may be small enough, however, that the CIs or significance tests will still be reasonably accurate.

An alternative method for constructing CIs that may hold promise is the Monte Carlo (MC) method of MacKinnon et al. (2004). In this approach, the sampling distribution for the effect of interest is not assumed to be normal and is instead simulated from the model estimates and their asymptotic variances and covariances (a form of parametric bootstrapping). For instance, to simulate the sampling distribution of the average indirect effect, one would define a multinormal distribution with means equal to $\hat{a}$, $\hat{b}$, and $\hat{\sigma}_{a_j,b_j}$ and covariance matrix equal to the estimated covariance matrix of these estimates. One would then take random values from this multinormal distribution and plug them into Equation 5 to compute the average indirect effect. Collecting the results over many draws provides a simulated sampling distribution for the average indirect effect. One would then obtain confidence limits for the average indirect effect by taking the corresponding percentiles of this simulated sampling distribution (e.g., 2.5th and 97.5th for a 95% CI). The advantage of this approach is that it does not assume normality for the sampling distributions of the average indirect and total effects.

Summary of Procedures

In conclusion, when one estimates lower level mediation models with random effects, interest may center on both the average causal effects and potential heterogeneity in these causal effects across the Level 2 units of the population. The average indirect and total effects can be computed via Equations 5 and 7. Estimates obtained from these equations indicate the strength of the indirect and total effects for an average Level 2 unit in the population. Allowing the indirect and total effects to be random, however, also implies that there is variability about these average values across Level 2 units. This heterogeneity in the strength of the indirect and total effects in the population can be characterized by the variance formulas given in Equations 6 and 8. Therefore, for instance, the variance obtained from Equation 6 indicates the extent to which the magnitude of the indirect effect differs across Level 2 units.

We have also suggested two possible ways to make inferences for the average indirect and total effect estimates. In the first approach, the standard errors of the average indirect and total effects are computed as the square roots of Equations 9 and 10. One can then use these standard errors to form CIs (or significance tests) by assuming that the estimates of the average indirect and total effects have normal sampling distributions, as shown in Equations 11 and 12. A second approach is to simulate the sampling distributions of the average indirect and total effects. This second approach is more difficult to implement, but it has the advantage that the simulated sampling distributions of the average indirect and total effect estimates will have the theoretically correct (nonnormal) forms. Both tests are implemented in a SAS macro available on the Web at http://dx.doi.org/10.1037/1082-989X.11.2.142.supp. Given the trade-offs involved, we now turn to a comparison of the performance of these two alternative procedures for making inferences about the average indirect and total effects.

Performance With Simulated Data

As a preliminary investigation of the accuracy of the estimates and CIs just described, we conducted a modest simulation study based on the example presented in Kenny et al. (2003, pp. 122–123). In their example, the random intercept for the M equation, d_{Mj}, had a mean of 0 and a variance of .6, and the random intercept for the Y equation, d_{Yj}, had a mean of zero and a variance of .4. These two random effects were normally distributed and were uncorrelated with each other and with the other random effects in the model. For the M equation, the Level 1 residual variance was set to $\sigma^2_{e_M} = .65$; for the Y equation, it was set to $\sigma^2_{e_Y} = .45$. The causal paths were simulated as follows: a_j and b_j were both normally distributed with means of $a = b = .6$ and variances of $\sigma^2_{a_j} = \sigma^2_{b_j} = .16$ and c'_j was normally distributed with a mean of $c' = .2$ and a variance of $\sigma^2_{c'_j} = .04$. The covariance between a_j and b_j was $\sigma_{a_j,b_j} = .113$, yielding a correlation of .706. Neither a_j nor b_j was correlated with c'_j. Last, the predictor X was simulated from the equation $X_{ij} = \bar{X}_j + e_{Xij}$, where $\bar{X}_j \sim N(0, 1)$ and $e_{Xij} \sim N(0, 1)$.

To generate our own simulated data, we used this model as a base from which we varied four design factors. In varying these factors, we focused specifically on how they might influence the average indirect effect estimate, as this effect is typically of most interest (results differed little for the average total effect). We chose the first two factors of the simulation study to manipulate the effect size of the average indirect effect as expressed in Equation 5. First, the magnitude of the a and b parameters was set to either $a = b = .3$ or $a = b = .6$. The higher values for a and b are identical to those used by Kenny et al. (2003), whereas the lower values are close to the smaller effect sizes considered by Krull and MacKinnon (2001) for similar models. The second design factor was the magnitude of the covariance between a_j and b_j. This parameter was set to one of three values, $\sigma_{a_j,b_j} = -.113$, 0, or .113 (correlations of $-.706$,

.000, and .706, respectively). The next two design factors manipulated the total sample size. We first set the number of Level 2 units to $N = 25$, 50, 100, or 200. We then set the number of observations per Level 2 unit to $n_j = 4$, 8, 16, or 32. These sample sizes are consistent with those used by Krull and MacKinnon (2001). The final design factor concerned the distributions of the random effects a_j and b_j. To evaluate the robustness of the estimates, we simulated a_j and b_j either from normal distributions or from $\chi^2(3)$ distributions, having skew 1.63 and kurtosis 4.

Together, these four factors combined to yield 144 conditions in a fully factorial design. In addition, to investigate Type I errors, we simulated supplemental data where $a = b = 0$ and $\sigma_{a_j,b_j} = 0$ at each sample size and for both distributional conditions, which resulted in an additional 24 conditions. Within each cell of the design, we simulated 500 samples of data, which resulted in a total of 84,000 replications (further details on the data generation are available from Daniel J. Bauer on request). We then fitted the model specified in Equation 4 to each data set using SAS PROC MIXED with the restricted maximum likelihood estimator and a maximum of 1,000 iterations.

We now consider four aspects of the results in turn. First, given the complexity of the model, we evaluate potential problems of estimation by examining the rate of model nonconvergence and boundary solutions (e.g., solutions for which a variance estimate is constrained to zero because it would otherwise become negative). Second, we consider under what conditions the estimate of the average indirect effect is appreciably biased. As part of this evaluation, we consider potential bias in the estimate of σ_{a_j,b_j} in particular, as we could not estimate this part of the average indirect effect directly using the previous two-step approach (but it is provided by the simultaneous modeling approach advocated here). Third, we examine Type I error rates for the null hypothesis test of the average indirect effect. Last, we compare the coverage rates and power of CIs constructed with the normal approximation and the MC method.

Difficulty of Estimation

The outcome of fitting each model was categorized as either nonconverged, a boundary solution, or an unconstrained (nonproblematic) solution. Boundary solutions were defined as solutions in which one or more random effect was constrained to have a variance of zero or the correlation between two random effects was constrained to be 1 to obtain a solution in which all estimates attained permissible values. Though constraints to boundary values do not invalidate the model results, they are typically seen as an undesirable result and suggest that the available data are not sufficient to support the complexity of the model. It is not surprising that the results shown in Table 1 indicate that as the sample size increased, by increased N or n_j, fewer replications resulted in nonconvergence and more replications produced unconstrained solutions. The two sample size factors did not have an equal influence, however. Given a finite number of total observations, the model becomes more stable if there are more observations per Level 2 unit at the expense of the number of Level 2 units. For instance, with 800 possible observations, setting $n_j = 16$ and $N = 50$ resulted in 92% unconstrained solutions, setting $n_j = 8$ and $N = 100$ resulted in 87% unconstrained solutions, and setting $n_j = 4$ and $N = 200$ resulted in 78% unconstrained solutions. In addition to these results, small effects of the other factors were also observed: The models were more likely to produce unconstrained solutions if all of the random effects were normally distributed, if the a and b values were small (.3), and if the covariance of a_j and b_j was zero. The latter effect can be understood as a consequence of the population value for the corresponding correlation being farther away from the boundary values of -1 and 1. Why smaller values of a and b produced more unconstrained solutions is less clear.

Bias of the Estimate

We next evaluated the bias of the average indirect effect estimate. Bias was computed as the difference between the mean estimate and the corresponding population value. The average indirect effect in the population was computed from the true parameter values and Equation 5. Across all 144 conditions in the factorial design, the mean bias for the average indirect effect estimate was $-.001$, with a minimum of $-.076$ and a maximum of .027. Further, the absolute bias was less than .01 in all but 18 conditions.[5] These 18 conditions included 14 conditions with nonnormal random effects and 9 conditions with the minimum total sample size of $N = 25$ and $n_j = 4$. A similar evaluation of the bias in the covariance estimate for a_j and b_j revealed that, across the 144 conditions, the mean bias was .0002, with a minimum of $-.033$ and a maximum of .041. The absolute bias was less than .01 in all but 12 conditions, all of which were conditions for which $n_j = 4$. Ten of the 12 conditions included nonnormal random effects, and 7 of the 12 conditions had the minimum number of Level 2 units ($N = 25$).

Overall, these results indicated that the estimate of the average indirect effect (and the covariance estimate for a_j

[5] We also evaluated the relative bias of the estimates but quickly ran into difficulty interpreting these values. That is, relative bias is computed as the ratio of bias to the true parameter value, but if the true parameter value is small even relatively trivial levels of bias will result in high relative bias. Consistent with this, we found that the relative bias was highest for conditions in which the population parameter value was lowest. For this reason, we focus in the text on bias as opposed to relative bias.

Table 1
Result of Fitting the Lower Level Mediation Model With Random Causal Effects as a Function of the Number of Level 2 Units (N) and the Number of Level 1 Units Within Each Level 2 Unit (n_j)

Observations	Result of model fitting (% of replications)		
	Failed to converge	Boundary solution	Unconstrained solution
	$N = 25$		
$n_j = 4$	93.03	4.90	2.07
$n_j = 8$	56.63	23.10	20.27
$n_j = 16$	16.18	23.02	60.80
	$N = 50$		
$n_j = 4$	54.70	26.93	18.37
$n_j = 8$	13.93	28.10	57.97
$n_j = 16$	1.18	6.72	92.10
	$N = 100$		
$n_j = 4$	16.52	34.13	49.35
$n_j = 8$	1.80	10.97	87.23
$n_j = 16$	0.37	0.67	98.97
	$N = 200$		
$n_j = 4$	2.50	19.52	77.98
$n_j = 8$	0.58	1.70	97.72
$n_j = 16$	0.27	0.00	99.73

Note. Results are collapsed over conditions that vary in the population parameter values and in the distributions of the random effects (normal or nonnormal).

and b_j) was largely unbiased except for a few of the nonnormal random effects or very small sample conditions. It is noteworthy that these are also the conditions for which it was difficult to estimate the model, which provides converging evidence that we need larger samples to obtain reliable results.

Type I Errors

Using the 24 supplementary cells of the simulation design for which the population value of the average indirect effect was zero, we next considered the Type I error rates for tests of the average indirect effect (setting the nominal error rate at 5%). We calculated 95% CIs using the normal approximation in Equation 11 and the MC method of MacKinnon et al. (2004). The latter method involved taking 50,000 random draws from the estimated sampling distribution of the estimates (i.e., the multivariate normal sampling distribution of $\hat{a}$, $\hat{b}$, and $\hat{\sigma}_{a_j,b_j}$), calculating Equation 5 for each draw, and then computing the 2.5th and 97.5th percentiles of the resulting values. Both the normal approximation and the MC method require an estimate of the covariance matrix of the fixed effects (e.g., *a* and *b*). Typically, one computes these values by taking the inverse of the information matrix for the estimates. For multilevel models, however, this procedure is known to underestimate the sampling variability of the fixed effect estimates because the variance components of the model are treated as known (when they are, in fact, estimated). Kackar and Harville (1984) provided a method for inflating the elements of the covariance matrix of the fixed effects to correct for this bias. We compare the CIs we obtained by using the default covariance matrix of the fixed effects versus the inflated Kackar–Harville (K-H) covariance matrix of the fixed effects. To do so, we calculated Type I error rates as the percentage of replications in which the 95% CI for the average indirect effect failed to cover zero.

Tables 2 and 3 compare the Type I error rates for the four types of CIs for normally and nonnormally distributed random effects, respectively. The results indicate that all four CIs tended to produce Type I error rates below the nominal level, especially at the lower sample sizes. CIs based on the MC method, however, reflected the nominal Type I error rate better than the other CIs. Further examination of the data indicated that, at the smallest sample sizes, the standard errors of the average indirect effect estimate were overestimated, producing CIs that were too wide and thus more likely to include zero. We now consider whether this is also the case when the null hypothesis is false.

Table 2
Type I Error Rates: Normal Random Effects

Condition	Replications[a]	Normal approximation		Monte Carlo method	
		Usual cov[b]	K-H cov[c]	Usual cov[b]	K-H cov[c]
		$N = 25$			
1. $n_j = 4$	70	0.00	0.00	0.00	0.00
2. $n_j = 8$	337	2.97	2.97	5.34	5.34
3. $n_j = 16$	484	2.48	2.48	4.34	4.13
		$N = 50$			
4. $n_j = 4$	351	4.27	4.27	4.84	4.56
5. $n_j = 8$	486	3.70	3.70	4.12	4.12
6. $n_j = 16$	500	3.20	3.20	4.20	4.20
		$N = 100$			
7. $n_j = 4$	482	4.98	4.98	5.39	5.39
8. $n_j = 8$	499	4.41	4.41	4.61	4.61
9. $n_j = 16$	500	6.00	6.00	6.20	6.00
		$N = 200$			
10. $n_j = 4$	500	4.00	4.00	4.00	4.00
11. $n_j = 8$	500	6.00	6.00	6.20	6.20
12. $n_j = 16$	500	4.80	4.60	5.00	4.60
		Collapsing			
1 through 12	5,209	4.24	4.22	4.86	4.76
2 through 12	5,139	4.30	4.28	4.92	4.83
4 through 12	4,318	4.61	4.59	4.96	4.86

Note. Results are presented by the number of Level 2 units (N) and the number of Level 1 units per Level 2 unit (n_j). The nominal Type I error rate was set at 5%.
[a] Number of replications resulting in a converged solution from which Type I error rates were calculated (out of 500, except for results collapsed over conditions). [b] Covariance matrix of a and b estimated by the inverse of the information matrix. [c] Covariance matrix of a and b estimated by the method of Kackar and Harville (1984).

CI Coverage Rates

Coverage rates for the CIs are presented in Table 4 by sample size and in Table 5 by the magnitude of the parameter values in the population model. Consistent with the results on Type I errors, the CIs were too wide in the smallest sample condition ($N = 25$, $n_j = 4$) when the random effects were normal. Outside of this condition, however, the CIs actually tended to be too narrow, producing slightly lower than 95% coverage of the true population parameter values. The dominant factor determining the coverage rates of the CIs was the distribution of the random effects. In general, regardless of the method for computing the CIs, the coverage rates hovered around 94% when the random effects were normally distributed and dropped to about 90% when the random effects were nonnormally distributed. The lower coverage rates for the nonnormal conditions reflected a general underestimation of the sampling variances of all of the estimates in the model.

Tables 4 and 5 also point to some clear differences between methods for computing the CIs. Across replications, the CIs constructed via Equation 11 performed slightly better than those computed via the MC method. This advantage of the normal approximation method was seen for 105 of 144 conditions (73%) but was stronger when the random effects were nonnormal. Additionally, Tables 4 and 5 show that the K-H covariance matrix of the fixed effects produced CIs with better coverage rates. Overall, we attained the best coverage rates by combining the normal approximation of Equation 11 with the K-H procedure. The MC method with the K-H procedure performed nearly as well.

Finally, Tables 4 and 5 indicate that the effect of nonnormal random effects on the coverage rates of the CIs was moderated by several factors. First, the coverage rates were slightly better at the smallest sample sizes (low N combined with low n_j). Second, the coverage rates were least affected by nonnormality when the covariance of the random effects

Table 3
Type I Error Rates: Nonnormal Random Effects

Condition	Replications[a]	Normal approximation		Monte Carlo method	
		Usual cov[b]	K-H cov[c]	Usual cov[b]	K-H cov[c]
		$N = 25$			
1. $n_j = 4$	22	0.00	0.00	0.00	0.00
2. $n_j = 8$	198	1.01	1.01	2.53	2.53
3. $n_j = 16$	435	2.07	2.07	3.45	3.45
		$N = 50$			
4. $n_j = 4$	205	1.46	1.46	2.44	2.44
5. $n_j = 8$	450	4.67	4.44	4.67	4.67
6. $n_j = 16$	495	4.65	4.65	5.25	5.25
		$N = 100$			
7. $n_j = 4$	428	5.14	5.14	5.61	5.61
8. $n_j = 8$	495	3.43	3.43	4.04	3.84
9. $n_j = 16$	500	4.20	4.20	4.80	4.80
		$N = 200$			
10. $n_j = 4$	497	4.83	4.63	5.03	5.03
11. $n_j = 8$	500	4.00	4.00	4.40	4.20
12. $n_j = 16$	500	4.60	4.60	5.00	5.00
		Collapsing			
1 through 12	4,725	3.92	3.87	4.49	4.44
5 through 12	3,865	4.42	4.37	4.84	4.79

Note. Results are presented by the number of Level 2 units (N) and the number of Level 1 units per Level 2 unit (n_j). The nominal Type I error rate was set at 5%.
[a] Number of replications resulting in a converged solution from which Type 1 error rates were calculated (out of 500, except for results collapsed over conditions). [b] Covariance matrix of a and b estimated by the inverse of the information matrix. [c] Covariance matrix of a and b estimated by the method of Kackar and Harville (1984).

was zero, because of bias in the estimate of the covariance when it was not zero. As would be expected, this effect was greatest when the a and b parameters were small (reflecting the then proportionally larger contribution of the covariance to the expected value). In contrast, these same factors minimally affected the coverage rates when the random effects were all normally distributed.

Power

To evaluate power, we calculated the proportion of replications with CIs for the average indirect effect excluding zero for each of the 144 cells in the factorial design. The results were consistent with the CI coverage rates. Because the K-H procedure necessarily results in larger estimates of sampling variability, this procedure produces wider CIs and, hence, lower power. Across conditions, however, the average loss in power due to the use of the K-H procedure never exceeded .0022 for either the normal approximation or the MC method. Similarly, because the MC method produced consistently narrower CIs than the normal approximation method, the MC method also resulted in superior power in 136 of the 144 conditions in the simulation study (94.4%). These differences in power were, however, also quite small, never differing by more than .0045.

Summary

In total, both the normal approximation method and the MC method for constructing CIs performed well when the random effects were normally distributed and the K-H procedure was used to estimate the sampling covariance matrix of the fixed effects. Between the two, the MC method was slightly more powerful and attained nominal Type I error rates at smaller sample sizes but also had slightly lower CI coverage rates. When the random effects were nonnormal, effect estimates remained unbiased in most cells of the design, but the CIs did not have accurate coverage rates. One exception was when the null hypothesis was true, in which case the CIs covered zero approximately 95% of the time, yielding roughly 5% Type I error rates, except in the smallest samples. We now turn to a slightly more compli-

Table 4
Confidence Interval Coverage by Distribution and Sample Size

Condition	Normally distributed a_j and b_j				Nonnormally distributed a_j and b_j			
	Normal approximation		Monte Carlo method		Normal approximation		Monte Carlo method	
	Usual cov[a]	K-H cov[b]	Usual cov[a]	K-H cov[b]	Usual cov[a]	K-H cov[b]	Usual cov[a]	K-H cov[b]
				$N = 25$				
$n_j = 4$	96.43	97.92	95.83	97.02	92.68	95.12	91.46	95.12
$n_j = 8$	93.75	94.55	93.62	94.43	92.71	93.82	91.89	93.31
$n_j = 16$	93.74	93.93	93.36	93.66	89.64	90.10	89.18	89.68
				$N = 50$				
$n_j = 4$	93.51	94.70	93.57	94.52	91.43	92.97	91.04	92.49
$n_j = 8$	94.12	94.26	94.01	94.37	90.90	91.46	90.16	90.82
$n_j = 16$	94.63	94.63	94.59	94.69	91.02	91.09	90.41	90.55
				$N = 100$				
$n_j = 4$	93.04	93.49	92.93	93.49	91.22	91.87	90.92	91.61
$n_j = 8$	93.85	93.98	93.88	93.98	90.13	90.37	89.85	90.02
$n_j = 16$	94.76	94.83	94.73	94.76	90.27	90.34	89.74	89.87
				$N = 200$				
$n_j = 4$	94.08	94.35	94.18	94.15	89.29	89.50	88.91	89.05
$n_j = 8$	94.63	94.63	94.33	94.46	90.67	90.77	90.64	90.74
$n_j = 16$	94.60	94.60	95.00	95.00	89.68	89.75	89.78	89.88
All replications	94.14	94.39	94.10	94.35	90.44	90.77	90.06	90.43

Note. Results are presented by the number of Level 2 units (N) and the number of Level 1 units per Level 2 unit (n_j).
[a] Covariance matrix of a and b estimated by the inverse of the information matrix. [b] Covariance matrix of a and b estimated by the method of Kackar and Harville (1984).

cated multilevel mediation model, one in which the strength of the causal effects depends on other predictors in the model.

Moderated Multilevel Mediation

Given evidence of significant random effects for either the a_j or b_j paths, one may wish to add factors to the model to explain this variability (Kenny et al., 2003). A simple version of this model, with a single Level 2 predictor W, is diagrammed in Figure 4. Significant prediction of either a_j or b_j by W would represent a case of moderated mediation, in which the strength of the indirect effect of the Level 1 predictor X depends on the Level 2 predictor W. The Level 2 predictor may also impact c'_j. Though this is not a case of moderated mediation as usually defined, significant prediction solely of c'_j still alters the balance of direct and indirect effects relative to the total effects and may thus be of substantive interest. The topic of moderated mediation (of several forms) has been explored more extensively for single-level regression models (e.g., James & Brett, 1984; Judd & Kenny, 1981; Lance, 1988; Morgan-Lopez, 2003; Muller, Judd, & Yzerbyt, 2005; Preacher, Rucker, & Hayes, 2006; Wegener & Fabrigar, 2000). We now draw on this literature as well as our preceding developments to propose a strategy for investigating such effects in multilevel models.

The moderated multilevel mediation model diagrammed in Figure 4 has the same Level 1 equations as before, namely,

$$M_{ij} = d_{Mj} + a_j X_{ij} + e_{Mij},$$

$$Y_{ij} = d_{Yj} + b_j M_{ij} + c'_j X_{ij} + e_{Yij} \quad (13)$$

but now each random coefficient is expressed as a linear function of W in the Level 2 model:

$$d_{M_j} = \gamma_{d_M 0} + \gamma_{d_M 1} W_j + u_{d_{Mj}}$$

$$a_j = \gamma_{a0} + \gamma_{a1} W_j + u_{aj}$$

$$d_{Y_j} = \gamma_{d_Y 0} + \gamma_{d_Y 1} W_j + u_{d_{Yj}}$$

$$b_j = \gamma_{b0} + \gamma_{b1} W_j + u_{bj}$$

$$c'_j = \gamma_{c'0} + \gamma_{c'1} W_j + u_{c'j}. \quad (14)$$

The assumption of normality previously made for a_j, b_j, and c'_j in Equation 2 now shifts to the residuals of the random coefficients, designated by the symbol u. In particular, these

Table 5
Confidence Interval Coverage by Distribution and Effect Size

Condition	Normally distributed a_j and b_j				Nonnormally distributed a_j and b_j			
	Normal approximation		Monte Carlo method		Normal approximation		Monte Carlo method	
	Usual cov[a]	K-H cov[b]	Usual cov[a]	K-H cov[b]	Usual cov[a]	K-H cov[b]	Usual cov[a]	K-H cov[b]
				$a = b = .3$				
$\sigma_{a_j,b_j} = -.113$	94.95	95.14	94.75	94.85	91.12	91.41	90.51	95.12
$\sigma_{a_j,b_j} = 0$	95.07	95.27	94.78	94.90	93.72	93.96	93.34	93.31
$\sigma_{a_j,b_j} = .113$	93.60	93.72	94.23	94.31	84.33	84.62	83.65	89.68
				$a = b = .6$				
$\sigma_{a_j,b_j} = -.113$	93.78	94.21	93.42	93.92	91.90	92.41	91.53	92.49
$\sigma_{a_j,b_j} = 0$	93.91	94.17	93.79	94.19	93.69	94.21	93.71	90.82
$\sigma_{a_j,b_j} = .113$	93.51	93.80	93.57	93.92	87.49	87.76	87.26	90.55
All replications	94.14	94.39	94.10	94.35	90.44	90.77	90.06	90.43

Note. Results are presented by the value of fixed effects in the population (a and b) and the covariance of the random effects in the population (σ_{a_j,b_j}). [a] Covariance matrix of a and b estimated by the inverse of the information matrix. [b] Covariance matrix of a and b estimated by the method of Kackar and Harville (1984).

residuals are assumed to be normal with an expected value of zero and a full covariance matrix. The symbol γ is reserved for the fixed effects of the model. Note that one may or may not wish to include W as a predictor of all of the random slopes (a_j, b_j, and c'_j). However, when W is included as a predictor of one of the random slopes for an equation, we strongly recommend that it be included as a predictor of the random intercept for the same equation.[6]

The conditional expected value of the indirect effect is

$$E(a_j b_j | W_j = w) = (\gamma_{a0} + \gamma_{a1} w)(\gamma_{b0} + \gamma_{b1} w) + \sigma_{u_{aj},u_{bj}}. \quad (15)$$

To test moderation of the indirect effect, we can either test γ_{a1} and γ_{b1} separately (the default of most software) or, perhaps preferably, perform a joint test of γ_{a1} and γ_{b1} (Raudenbush & Bryk, 2002, pp. 58–61). Note that this joint test is only possible with the simultaneous estimation strategy advocated in this article. Regardless of which testing strategy is taken, rejection of the null hypothesis would provide evidence of moderated mediation.

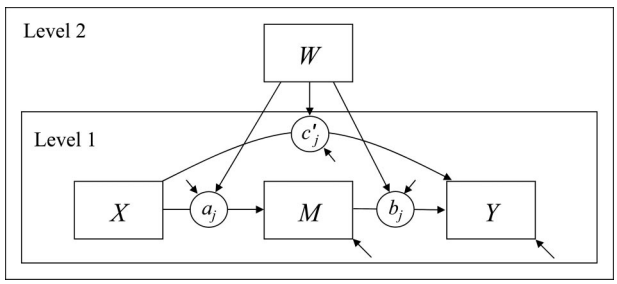

Figure 4. Moderated lower level mediation model: Nested frames indicate levels of sampling, boxes indicate variables, arrows without circles represent fixed effects, arrows with circles represent random effects, and arrows not originating from variables indicate residuals. Note that the arrows from W to the random effects indicate moderation of the hypothesized causal effects. Predictions of M and Y by W are not shown on the diagram but would be included in the model.

Given evidence of moderated mediation, we may wish to evaluate the simple indirect effects, that is, the value of Equation 15 at various levels of W. Inference tests of the simple indirect effects are also possible. Note that in the special case in which $W = 0$, the simple indirect effect obtained from Equation 15 is simply

$$E(a_j b_j | W_j = 0) = \gamma_{a0}\gamma_{b0} + \sigma_{u_{aj},u_{bj}}. \quad (16)$$

In this special case, Equation 16 is of the same basic form as Equation 5, and we can calculate the sampling variance of the simple indirect effect with the analogue to Equation 9, or

$$\begin{aligned}\mathrm{Var}(\hat{\gamma}_{a0}\hat{\gamma}_{b0} + \hat{\sigma}_{u_{aj},u_{bj}}) = {} & \gamma_{b0}^2 \mathrm{Var}(\hat{\gamma}_{a0}) + \gamma_{a0}^2 \mathrm{Var}(\gamma_{b0}) \\ & + \mathrm{Var}(\hat{\gamma}_{a0})\mathrm{Var}(\hat{\gamma}_{b0}) + 2\gamma_{a0}\gamma_{b0}\mathrm{Cov}(\hat{\gamma}_{a0},\hat{\gamma}_{b0}) \\ & + \mathrm{Cov}(\hat{\gamma}_{a0},\hat{\gamma}_{b0})^2 + \mathrm{Var}(\hat{\sigma}_{u_{aj},u_{bj}}). \quad (17)\end{aligned}$$

[6] When a Level 2 covariate predicts a random slope, this results in a cross-level interaction. Likewise, when a Level 2 covariate predicts a random intercept, this results in a main effect. One can see this by forming the reduced form equation for the model (i.e., by substituting the Level 2 equations into the Level 1 equations). Typically, one should always include main effects when evaluating interactions. See Bauer and Curran (2005) for additional discussion of these issues for multilevel models.

Replacing the population values in Equation 17 with our sample estimates, we can use the normal approximation or MC method to construct a CI for the simple indirect effect at $W = 0$.

Given these simplifications, one way to conduct tests of simple indirect effects at other levels of W is to rescale W so that the value of interest is equal to zero, much as is often done to evaluate other types of simple effects (e.g., see Aiken & West, 1991). For example, suppose w is the conditional value of W that is of interest. We can form a new variable $W_{CV} = W - w$ that is centered on this value. If we reestimate the model, replacing W with W_{CV}, the coefficients for the fixed effects will adjust to account for the change in scale, but the model will be equivalent. With these new coefficients, we can now apply Equations 16 and 17 to obtain the estimate and sampling variance of the simple indirect effect at $W = w$ (or $W_{CV} = 0$). In this way, we can compute a CI for the simple indirect effect at $W = w$. The simulation study conducted in the prior section suggests that these CIs will be reasonably accurate when the distributions of the random effects are fairly normal but less so when these distributions are nonnormal.

Similarly, the simple direct effect is

$$E(c_j'|W_j = w) = \gamma_{c0'} + \gamma_{c1'}w. \tag{18}$$

Here, the test of moderation is carried by a single parameter estimate, $\hat{\gamma}_{c1'}$. If this estimate is significantly different from zero, we may wish to probe the simple direct effects of X on Y. Again, by forming a new variable W_{CV}, centered at a conditional value of interest for W, we can obtain tests of the simple direct effects through the test of $\hat{\gamma}_{c0'}$. The simple total effects of X on Y are the sum of the simple indirect and direct effects, and inference tests can be conducted on the summed estimates in a similar fashion. More complex methods for probing moderation effects in multilevel models, including the calculation of regions of significance, are given in Bauer and Curran (2005) and could also be extended to moderated mediation models (Preacher et al., 2006).

Empirical Example

We now implement the procedures we have outlined with data from a study concerning perceptions of pain and stress in African American adolescents and adults with SCD (Gil et al., 2003, 2004). As noted in the introduction, multilevel models are applicable both to the case of individuals nested within groups and to the case of repeated measures nested within individuals. In this application, the latter structure is present. The 94 participants first completed a baseline interview and were then asked to complete a daily diary each evening for up to 6.5 months. The number of days diaries were completed ranged from 2 to 196, with a median of 69 days (a 75% completion rate). Diaries included three 100-mm visual analogue scales designed to measure the average physical sensation of pain (*PHYS*), emotional discomfort of the pain (*EMOT*), and level of stress (*STRESS*) that the participant experienced over the day. Scores were scaled to range from 0 to 10, with higher values indicating greater pain or stress.

Given the study design, the Level 1 units are the daily reports of pain and stress by the participants, and the Level 2 units are the participants. The model under investigation posits that the experience of physical pain will increase stress but that this effect will be largely mediated by the emotional response to the pain, that is, its perceived unpleasantness. Further, given evidence of individual differences in the direct and indirect effects of physical pain, we seek to explain these differences on the basis of Level 2 covariates collected at the baseline interview (i.e., extending to a moderated mediation model). To simplify our analysis of the data, the model we fit assumes that missing data, including data that are missing due to attrition, are missing at random (Raudenbush & Bryk, 2002, pp. 199–200). In actuality this may be untrue, and a more complex analysis may be required (Schafer & Graham, 2002, provided a review of current approaches to modeling nonignorably missing data). Finally, we used the K-H method for estimating the covariance matrix of the fixed effects and the MC method for estimating 95% CIs for the average indirect and total effects, although in all cases the inferences made would have been identical if we had used the normal approximation method. SAS code for implementing the models presented here is listed on the Web at http://dx.doi.org/10.1037/1082-989X.11.2.142.supp.

For our initial model, the Level 1 equations were specified as

$$\begin{aligned} \text{Emot}_{ij} &= d_{Ej} + a_j\text{Phys}_{ij} + e_{Eij} \\ \text{Stress}_{ij} &= d_{Sj} + b_j\text{Emot}_{ij} + c_j'\text{Phys}_{ij} + e_{Sij} \end{aligned} \tag{19}$$

where the subscript i references the repeated assessment and the subscript j references the participant. The coefficients a_j, b_j, and c_j' were all also allowed to be random, and all of the random effects were allowed to covary. Homogeneity of variance of the Level 1 residuals was assumed within each equation, but the residual variances were allowed to differ across equations, as shown here:

$$\begin{aligned} \text{Var}(e_{Eij}) &= \sigma_E^2 \\ \text{Var}(e_{Sij}) &= \sigma_S^2 . \end{aligned} \tag{20}$$

Aside from the random effects of the model, serial autocorrelation is a second common source of dependence in daily diary data (Schwartz & Stone, 1998; West & Hepworth, 1991). That is, ratings made close in time to one another are typically more highly correlated than ratings made farther apart in time. Although a variety of methods can be used to

account for serial autocorrelation (Beal & Weiss, 2003; West & Hepworth, 1991), in the present case we assumed a continuous-time autoregressive structure for both outcomes in Equation 19 following the recommendation of Schwartz and Stone (1998). This error structure assumes that the correlation between the residuals across assessments declines as an exponential function of the time lag between assessments and allows for varying time intervals between assessments. More formally, for any two observations i and i', the covariance between the residuals was specified as

$$\begin{aligned} \text{Cov}(e_{Eij}, e_{Ei'j}) &= \sigma_E^2 \rho_E^{d_{ii'}} \\ \text{Cov}(e_{Sij}, e_{Si'j}) &= \sigma_S^2 \rho_S^{d_{ii'}}, \end{aligned} \tag{21}$$

where $d_{ii'}$ is the number of days elapsed between observations i and i', ρ_E is the autoregressive parameter for emotional reactions to pain, and ρ_S is the autoregressive parameter for stress ratings. We estimated the full model with restricted maximum likelihood using the specification strategy detailed earlier in the article.

The results for the initial model are reported in Table 6. All three causal paths, a_j, b_j, and c'_j, varied significantly across persons and, on average, were positive and significantly different from zero. To better understand this pattern of effects, we next estimated the expected value and variance of the random indirect and total effects, $a_j b_j$ and $a_j b_j + c'_j$, respectively. The formulas for the variances require normality of the random effects to be exact, so we first checked this assumption by plotting the distributions of the empirical Bayes estimates for the random coefficients. We noted some nonnormality, as depicted in Figure 5 for the a_j and b_j estimates.

Next, noting that the covariance between a_j and b_j was estimated as .01 (r = .26), we determined the estimated average indirect effect to be .29 (95% CI = .16, .41) and the estimated average total effect to be .43 (95% CI = .36, .51). Given the nonnormality of the random effects, the results of our simulation study suggest that these CIs may be too narrow and should be interpreted with caution. In contrast, the effect estimates are probably unbiased, indicating that, on average, about 67% of the total effect of physical pain on stress was indirect (mediated by the emotional reaction to pain). Individual differences in the indirect effect were characterized by Equation 6 to have a variance of .11 (SD = .33). The variance of the total effect across participants was estimated from Equation 8 to be .08 (SD = .29). Although it may at first appear odd that the variance of the total effect was less than the variance of the indirect effect, the reason for this finding is that the direct effect was negatively correlated with both components of the indirect effect. That is, as the indirect effect went up, the direct effect went down, and vice versa, resulting in less variability in the total effect than in the indirect effect.

Table 6
Empirical Analysis Results

Effect	Estimate	*SE*	95% confidence limits	
			Lower	Upper
Fixed (average) effects				
d_E	0.028	0.026	−0.025	0.081
d_S	1.050	0.013	0.794	1.306
a	0.909	0.014	0.880	0.937
b	0.302	0.069	0.163	0.441
c'	0.148	0.059	0.029	0.267
Level 1 residual structure				
σ_E^2	0.206	0.003	0.199	0.212
σ_S^2	2.501	0.046	2.417	2.596
ρ_E	0.059	0.014	0.031	0.087
ρ_S	0.296	0.012	0.272	0.320

	Covariance/correlation matrix of random effects[a]				
	1	2	3	4	5
1. d_{E_j}	0.050	0.275	−0.276	−0.140	0.321
2. d_{S_j}	0.070	1.319	−0.040	−0.346	0.104
3. a_j	−0.007	−0.005	0.014	0.258	−0.228
4. b_j	−0.011	−0.137	0.011	0.119	−0.495
5. c'_j	0.016	0.026	−0.006	−0.037	0.048

[a] The variances of the random effects are shown on the diagonal, the covariances of the random effects are shown below the diagonal, and the correlations among the random effects are shown above the diagonal.

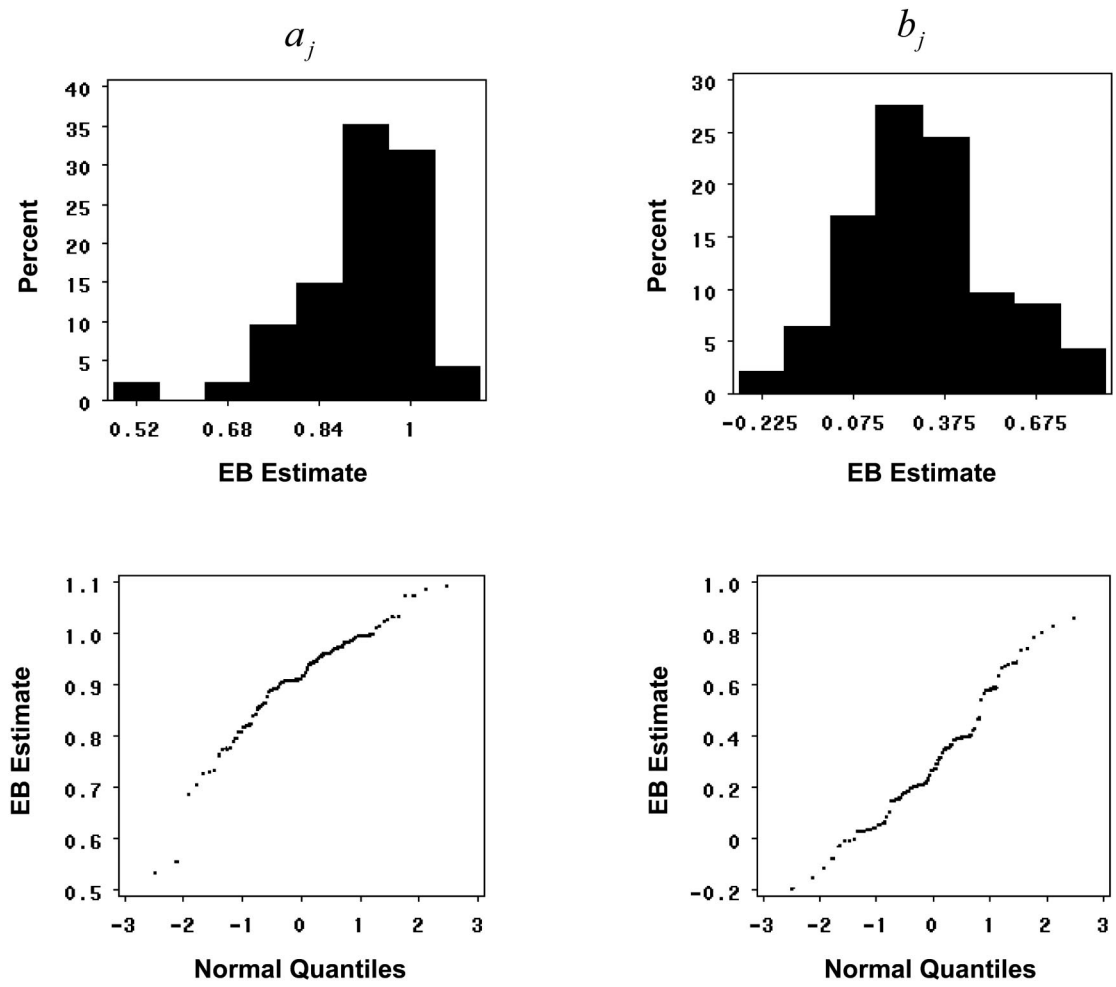

Figure 5. Diagnostic plots for judging the normality assumption for the random slopes involved in the indirect effect estimate. Top panels are histograms for the empirical Bayes (EB) estimates of the random effects obtained by fitting the initial model to the sickle cell disease data; bottom panels are normal quantile–quantile plots for the EB estimates.

To better understand individual differences in the strength of the direct, indirect, and total effects, we next extended the model we have described by adding a Level 2 predictor. The predictor of interest was the number of major acute complications of SCD (*ACUTE*) the patient reported having experienced within the past year. This predictor was added to the equation for each random effect. Of greatest interest was whether acute complications of SCD would moderate the indirect effect, or $a_j b_j$. The coefficient estimates indicated that higher levels of acute complications reduced the role of emotional reactions to pain as a mediator between physical pain and stress; however, the simultaneous test of these coefficients was not significant, $\chi^2(2) = 4.75$, $p = .093$. In contrast, higher levels of acute complications appeared to increase the direct effect of physical pain on stress, but this was also not significant, $t(24.4) = 1.82$, $p = .080$.

Given that the moderating effects of *ACUTE* on the three causal paths were nonsignificant, the analysis would typically end here. However, simply to demonstrate the methodology described in the preceding section, we proceeded to probe the simple direct, indirect, and total effects of physical pain on stress at each observed count of acute complications, zero through six. As can be seen in Figure 6, although the total effect was relatively unaffected by the moderator, the balance of the direct and indirect effects shifted markedly. For participants reporting zero acute SCD complications in the past year, the effect of physical pain on stress was essentially all indirect. As the number of acute SCD complications reported by the participant rose, however, this indirect effect diminished to nearly zero, and the direct effect of physical pain on stress increased to higher levels. Inferential tests of the simple direct and indirect effects revealed significant indirect effects for participants

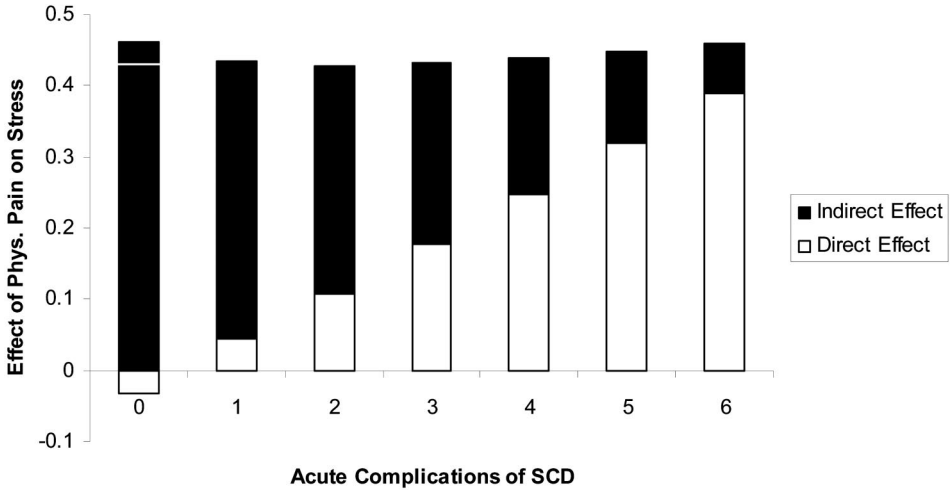

Figure 6. Bars indicate the decomposition of the average causal effect of physical pain on stress into a direct effect and an indirect effect through emotional reactions to pain. The total height of each column conveys the magnitude of the total effect, with the exception of the first column, where the direct effect is negative and the total effect is indicated with a white horizontal line. Note that the total effect of physical pain on stress is relatively constant but that the mediation of this effect by emotional reactions to pain wanes as the severity of sickle cell disease (SCD) increases.

reporting four or fewer acute SCD complications in the past year and significant direct effects for participants reporting three or more complications. The simple total effects were relatively stable and statistically significant across all observed levels of acute SCD complications.

Epistemological Issues in Assessing Mediation

Having established and demonstrated procedures for evaluating hypotheses of lower level mediation in multilevel models, we now review several important epistemological issues that must be considered whenever mediation is to be assessed. The foremost assumption of any mediation model is that the distal predictor and the mediator exert causal effects on their respective dependent variables. Necessary preconditions for causal inference are that (a) the variables involved must covary with each other, (b) causes must occur before their presumed effects, and (c) sources of spurious covariation should be eliminated (Frazier, Tix, & Barron, 2004). To satisfy the second condition, *X* should be measured prior to *M* and *M* should be measured prior to *Y* whenever possible (Gollob & Reichardt, 1987, 1991). Experimental manipulation of *X* can greatly strengthen the causal inference, as can secondary manipulation of *M* (Spencer, Zanna, & Fong, 2005). With respect to the third condition, MacKinnon et al. (2002) and Holland (1988) noted that the independence assumption made for the cross-equation residuals in Equation 1 is particularly questionable. Violation of this assumption can occur in a number of ways, including the omission of important variables, the presence of common method variance, and model misspecification. The result may be bias in the indirect effect estimate. Confidence in the independence assumption can be increased through the inclusion of potential common causes of *Y* and *M* (other than *X*) in the model and through the use of multiple methods of data collection. Further recommendations for improving causal inferences can be found in Berk (1988, 2004), Cole and Maxwell (2003), Gollob and Reichardt (1987, 1991), and Hoyle and Robinson (2004).

With regard to model specification, one issue that is unique to the lower level mediation model is the partitioning of the covariance structure for the Level 1 observations. If the Level 1 residual covariance structure is not correctly specified, then this may lead one to overestimate the variance components for the random effects. As a consequence, one could conclude that there is more heterogeneity in the hypothesized causal effects of the model than is in fact present in the population of Level 2 units. A second specification issue unique to the multilevel setting is which effects should have random versus fixed components. This issue is important because misspecification of the random effects could also bias the indirect effect estimate. Ideally, the theoretical model of the causal processes will indicate which of the causal effects may show heterogeneity across Level 2 units. If, however, theory does not provide strong guidance, then researchers may wish to empirically evaluate the issue by including random effects for each causal path and assessing the magnitude of the estimated variance components for these effects.

In our view, it is rarely the case that an investigator is able to address all of the aforementioned issues simultaneously. For instance, for the SCD analyses we have presented, the

measures are contemporaneous, and questions could be raised about whether the causal effects are in the correct direction (or bidirectional), whether the estimates are inflated by common method variance, and whether an autoregressive structure was sufficient to capture the over-time residual correlations. Like the SCD analyses, definitive evidence of mediation is often not obtained from a single model; instead, multiple studies are typically necessary.

Limitations and Future Directions

Although we believe that the simultaneous approach we have proposed to modeling random indirect effects in $1 \rightarrow 1 \rightarrow 1$ multilevel mediation models has a number of strengths, we would be remiss if we did not also note its limitations. First, our simulation study considers only a small set of the possible conditions that might be encountered in practice. Indeed, our own analysis of the SCD data involves unbalanced data, potentially nonrandom attrition, and serial dependence, none of which is considered in the simulation study. We thus view our results as a preliminary indication of when and where the performance of the proposed procedures can be expected to suffer when the assumptions of the model are unmet, but more studies are needed to make more definitive conclusions.

One clear implication of our simulation study is that additional thought must be given to the best way to accommodate nonnormal random effects. One possibility is to retain the current model but implement a nonparametric bootstrap. Recently, Pituch, Stapleton, and Kang (in press) evaluated bootstrapping for conducting tests of mediation in multilevel models and obtained promising results. Their research was, however, restricted to models with normally distributed random effects and without random indirect effects. Unfortunately, the computational burden involved in bootstrapping the more complex models we consider may be prohibitive. For instance, the initial model we fitted to the SCD data took 40 min to estimate. Running the same model on 1,000 resampled data sets would thus take about 28 days of computing time. Some way must be found to improve the efficiency of the process if bootstrapping is to be a feasible option for these models.

An additional limitation of the present approach is that the model requires that the residuals for the mediator and the distal outcome be uncorrelated, an assumption that may not hold for reasons discussed in the prior section. For single-level models, one can potentially address this assumption by using a structural equation model (SEM) in which *M* and *Y* are latent variables and the residuals of the indicators for the latent variables covary across constructs (see, e.g., Bollen, 1989, p. 324). SEMs also account for potential measurement error in the observed variables and can be used to partial out common method variance (Bollen & Paxton, 1998; Kenny & Zautra, 2001). Despite many recent advances in multilevel structural equation modeling, however, it is currently infeasible to estimate a $1 \rightarrow 1 \rightarrow 1$ mediation model with random direct and indirect effects among the latent variables. The principal difficulty is that the typical estimators for multilevel SEMs allow for random intercepts but not random slopes (Bentler & Liang, 2003; Goldstein & MacDonald, 1988; B. O. Muthén, 1994; B. O. Muthén & Satorra, 1995). In principle, maximum likelihood with numerical integration (L. K. Muthén & Muthén, 2004) or Bayesian estimation techniques (Ansari, Jedidi, & Dube, 2002) can be used to include random slopes in a multilevel SEM. Future research should explore the potential use of these new methods with $1 \rightarrow 1 \rightarrow 1$ models for latent variables.

A number of practical difficulties may also be encountered by applied researchers wishing to fit these models. Foremost, in prototypical form, the model includes two random intercepts and three random slopes. In general, the estimation of a multilevel model becomes more difficult and computationally demanding as the number of random effects increases, and the present case is no exception. The identification of the variance components depends heavily on the number of Level 1 observations per Level 2 unit, whereas the accuracy with which they are estimated depends on the number of Level 2 units (Hox, 2002). In our simulation study, we encountered serious difficulty estimating the model when the number of Level 1 observations was small (e.g., four). Given this, certain kinds of study designs are more likely to permit the estimation of this model than others. For instance, studies using ecological momentary assessment (e.g., diary studies) typically yield many observations per unit and hence may be ideally suited for the estimation of lower level mediation models. In contrast, studies with fewer observations per unit may not support the estimation of five random effects, forcing the investigator to simplify the model by removing one or more random effects.

Conclusions

The ability to investigate heterogeneity in causal effects is one of the most attractive features of multilevel models. Most approaches for investigating mediation in multilevel models have, however, exclusively focused on fixed causal effects. In an important next step, Kenny et al. (2003) proposed an approach to evaluating $1 \rightarrow 1 \rightarrow 1$ mediation when the causal paths are random. In this model, the direct, indirect, and total effects vary in strength across the Level 2 units of the population. The present article extends the work of Kenny et al. (2003) in two important ways. First, we present a simultaneous modeling approach that provides all of the necessary information to calculate the variances of the random indirect and total effects as well as the standard errors of the average indirect and total effects. We believe that the formulas for calculating these standard errors are new results. Additionally, this simultaneous modeling ap-

proach offers the opportunity to conduct tests of moderated mediation when the components of the random indirect effect are predicted by a Level 2 variable. Second, we provide an initial study of the robustness of the estimates and CIs for the average indirect effect using simulation methodology. This study indicates that the estimate of the average indirect effect was unbiased under most conditions. The CIs also provided fairly accurate coverage rates when the random effects were normal, although they were too narrow when the random effects were nonnormal. With this caveat in mind, we believe that the developments we have presented offer applied researchers an improved approach for estimating indirect effects in $1 \rightarrow 1 \rightarrow 1$ multilevel mediation models. Future developments such as those we have suggested may offer further improvements, enabling applied researchers to more fully address hypotheses of mediation in multilevel data structures.

References

Aiken, L. S., & West, S. G. (1991). *Multiple regression: Testing and interpreting interactions.* Newbury Park, CA: Sage.

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

Ansari, A., Jedidi, K., & Dube, L. (2002). Heterogeneous factor analysis models: A Bayesian approach. *Psychometrika, 67,* 49–78.

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.

Bauer, D. J. (2003). Estimating multilevel linear models as structural equation models. *Journal of Educational and Behavioral Statistics, 28,* 135–167.

Bauer, D. J., & Curran, P. J. (2005). Probing interactions in fixed and multilevel regression: Inferential and graphical techniques. *Multivariate Behavioral Research, 40,* 373–400.

Beal, D. J., & Weiss, H. M. (2003). Methods of ecological momentary assessment in organizational research. *Organizational Research Methods, 6,* 440–464.

Bentler, P. M., & Liang, J. (2003). Two-level mean and covariance structures: Maximum likelihood via an EM algorithm. In S. P. Reise & N. Duan (Eds.), *Multilevel modeling: Methodological advances, issues, and applications* (pp. 53–70). Mahwah, NJ: Erlbaum.

Berk, R. A. (1988). Causal inference for sociological data. In N. J. Smelser (Ed.), *Handbook of sociology* (pp. 155–172). Newbury Park, CA: Sage.

Berk, R. A. (2004). *Regression analysis: A constructive critique.* Thousand Oaks, CA: Sage.

Bollen, K. A. (1987). Total, direct, and indirect effects in structural equation models. *Sociological Methodology 1987, 17,* 37–69.

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

Bollen, K. A., & Paxton, P. (1998). Detection and determinants of bias in subjective measures. *American Sociological Review, 63,* 465–478.

Catanzaro, S. J., & Laurent, J. (2004). Perceived family support, negative mood regulation expectancies, coping, and adolescent alcohol use: Evidence of mediation and moderation effects. *Addictive Behaviors, 29,* 1779–1797.

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

Demidenko, E. (2004). *Mixed models: Theory and applications.* Hoboken, NJ: Wiley.

Frazier, P. A., Tix, A. P., & Barron, K. E. (2004). Testing moderator and mediator effects in counseling psychology research. *Journal of Counseling Psychology, 51,* 115–134.

Gil, K. M., Carson, J. W., Porter, L. S., Ready, J., Valrie, C., Redding-Lallinger, R., & Daeschner, C. (2003). Daily stress and mood and their association with pain, health-care use and school activity in adolescents with sickle cell disease. *Journal of Pediatric Psychology, 28,* 363–373.

Gil, K. M., Carson, J. W., Porter, L. S., Scipio, C., Bediako, S. M., & Orringer, E. (2004). Daily mood and stress predict pain, health care use, and work activity in African American adults with sickle-cell disease. *Health Psychology, 23,* 267–274.

Goldstein, H. I., & McDonald, R. P. (1988). A general model for the analysis of multilevel data. *Psychometrika, 53,* 455–467.

Gollob, H. F., & Reichardt, C. S. (1987). Taking account of time lags in causal models. *Child Development, 58,* 80–92.

Gollob, H. F., & Reichardt, C. S. (1991). Interpreting and estimating indirect effects assuming time lags really matter. In L. M. Collins & J. L. Horn (Eds.), *Best methods for the analysis of change* (pp. 243–259). Washington, DC: American Psychological Association.

Goodman, L. A. (1960). On the exact variance of products. *Journal of the American Statistical Association, 55,* 708–713.

Holland, P. W. (1988). Causal inference, path analysis, and recursive structural equations models. *Sociological Methodology 1988, 18,* 449–484.

Hox, J. (2002). *Multilevel analysis: Techniques and applications.* Mahwah, NJ: Erlbaum.

Hoyle, R. H., & Robinson, J. C. (2004). Mediated and moderated effects in social psychological research: Measurement, design, and analysis issues. In C. Sansone, C. C. Morf, & A. T. Panter (Eds.), *The Sage handbook of methods in social psychology* (pp. 213–233). Thousand Oaks, CA: Sage.

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

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

Kackar, R. N., & Harville, D. A. (1984). Approximations for standard errors of estimators of fixed and random effects in

mixed linear models. *Journal of the American Statistical Association, 79,* 853–862.

Kendall, M. G., & Stuart, A. (1969). *The advanced theory of statistics* (3rd ed., Vol. 2). London: Charles Griffen & Co.

Kenny, D. A., Kashy, D. A., & Bolger, N. (1998). Data analysis in social psychology. In D. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *The handbook of social psychology* (4th ed., Vol. 1, pp. 223–265). New York: McGraw-Hill.

Kenny, D. A., Korchmaros, J. D., & Bolger, N. (2003). Lower level mediation in multilevel models. *Psychological Methods, 8,* 115–128.

Kenny, D. A., & Zautra, A. (2001). Trait-state models for longitudinal data. In L. M. Collins & A. G. Sayer (Eds.), *New methods for the analysis of change* (pp. 243–269). Washington, DC: American Psychological Association.

Kreft, I., & de Leeuw, J. (1998). *Introduction to multilevel modeling.* London: Sage.

Krull, J. L., & MacKinnon, D. P. (1999). Multilevel mediation modeling in group-based intervention studies. *Evaluation Review, 23,* 418–444.

Krull, J. L., & MacKinnon, D. P. (2001). Multilevel modeling of individual and group level mediated effects. *Multivariate Behavioral Research, 36,* 249–277.

Lance, C. E. (1988). Residual centering, exploratory and confirmatory moderator analysis, and decomposition of effects in path models containing interactions. *Applied Psychological Measurement, 12,* 163–175.

MacCallum, R. C., Kim, C., Malarkey, W., & Kiecolt-Glaser, J. (1997). Studying multivariate change using multilevel models and latent curve models. *Multivariate Behavioral Research, 32,* 215–253.

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

MacKinnon, D. P., Lockwood, C. M., Hoffman, J. M., West, S. G., & Sheets, V. (2002). A comparison of methods to test mediation and other intervening variable effects. *Psychological Methods, 7,* 83–104.

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

Morgan-Lopez, A. A. (2003). *A simulation study of the mediated baseline by treatment interaction effect in preventive intervention trials.* Unpublished doctoral dissertation, Arizona State University.

Muller, D., Judd, C. M., & Yzerbyt, V. Y. (2005). When moderation is mediated and mediation is moderated. *Journal of Personality and Social Psychology, 89,* 852–863.

Muthén, B. O. (1994). Multilevel covariance structure analysis. *Sociological Methods & Research, 22,* 376–398.

Muthén, B. O., & Satorra, A. (1995). Complex sample data in structural equation modeling. *Sociological Methodology, 1995,* 216–316.

Muthén, L. K., & Muthén, B. O. (2004). *Mplus user's guide* (3rd ed.). Los Angeles: Muthén & Muthén.

Pituch, K. A., Stapleton, L. M. & Kang, J. Y. (in press). A comparison of single sample and bootstrap methods to assess mediation in cluster randomized trials. *Multivariate Behavioral Research.*

Pituch, K. A., Whittaker, T. A., & Stapleton, L. M. (2005). A comparison of methods to test for mediation in multisite experiments. *Multivariate Behavioral Research, 40,* 1–24.

Preacher, K. J., Rucker, D. D., & Hayes, A. F. (2006). *Suggested procedures for addressing moderated mediation hypotheses.* Manuscript submitted for publication.

Rao, C. R. (1965). *Linear statistical inference and its applications.* New York: Wiley.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Newbury Park, CA: Sage.

Raudenbush, S. W., & Sampson, R. (1999). Assessing direct and indirect effects in multilevel designs with latent variables. *Sociological Methods & Research, 28,* 123–153.

Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods, 7,* 147–177.

Schwartz, J. E., & Stone, A. A. (1998). Strategies for analyzing ecological momentary assessment data. *Health Psychology, 17,* 6–16.

Shrout, P. E., & Bolger, N. (2002). Mediation in experimental and nonexperimental studies: New procedures and recommendations. *Psychological Methods, 7,* 422–445.

Spencer, S. J., Zanna, M. P., & Fong, G. T. (2005). Establishing a causal chain: Why experiments are often more effective than mediational analyses in examining psychological processes. *Journal of Personality and Social Psychology, 89,* 845–851.

Wegener, D. T., & Fabrigar, L. R. (2000). Analysis and design for nonexperimental data: Addressing causal and noncausal hypotheses. In H. T. Reis & C. M. Judd (Eds.), *Handbook of research methods in social and personality psychology* (pp. 412–450). Cambridge, England: Cambridge University Press.

Wei, M., Mallinckrodt, B., Russell, D. W., & Abraham, W. T. (2004). Maladaptive perfectionism as a mediator and moderator between adult attachment and depressive mood. *Journal of Counseling Psychology, 51,* 201–212.

West, S. G., & Hepworth, J. T. (1991). Statistical issues in the study of temporal data: Daily experiences. *Journal of Personality, 59,* 609–662.

(*Appendix follows on next page*)

Appendix

Derivation of Variance Estimates

This appendix addresses the derivation of four variances mentioned in the text:

1. variance of the indirect effect among Level 2 units (Equation 6),
2. variance of the total effect among Level 2 units (Equation 8),
3. sampling variance of the mean estimated indirect effect (Equation 9), and
4. sampling variance of the mean estimated total effect (Equation 10).

All derivations that follow rely on a large statistics literature stretching back to the 1930s addressing the distribution of the product of two random variables. The variances may be derived in a number of ways; for the sake of brevity, we rely on proofs using a second-order elaboration of the well-known multivariate delta method (Rao, 1965) for deriving the variance of an approximated function of variables. The idea behind this method is to expand the function of interest using a second-order Taylor series around the population means of the variables involved and then to find the variance of this expression. Under the assumption of normality, this method yields an exact result for the variances needed here.

Heterogeneity of the Indirect Effect Among Level 2 Units (Equation 6)

Kenny et al. (2003, Equation 11) quantified the variability of this indirect effect in multilevel modeling as

$$\mathrm{Var}(a_j b_j) = b^2\sigma_a^2 + a^2\sigma_b^2 + \sigma_a^2\sigma_b^2 + 2ab\sigma_{a_j,b_j} + \sigma_{a_j,b_j}^2. \quad \text{(A1)}$$

In what follows, we use a second-order elaboration of the multivariate delta method for deriving the variance of a function of random variables. In compact matrix notation,

$$\mathrm{Var}(g) = \mathbf{D}'\Sigma(\theta)\mathbf{D} + \tfrac{1}{2}tr\{(\mathbf{H}\Sigma(\theta))^2\}, \quad \text{(A2)}$$

where θ is a vector of the variable elements of g, $\mathbf{D}$ is the gradient (vector of first derivatives) of g with respect to θ evaluated at the means of a_j and b_j, $\mathbf{H}$ is the Hessian (matrix of second derivatives) of g with respect to θ, and $\Sigma(\theta)$ is the covariance matrix of θ. When $g = a_j b_j$, $(\delta g/\delta a)|_{a,b} = b$ and $(\delta g/\delta b)|_{a,b} = a$, and hence $\mathbf{D}' = [b\ a]$. Here,

$$\Sigma(\theta) = \begin{bmatrix} \sigma_{a_j}^2 & \sigma_{a_j,b_j} \\ \sigma_{a_j,b_j} & \sigma_{b_j}^2 \end{bmatrix} \quad \text{(A3)}$$

and

$$\mathbf{H} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad \text{(A4)}$$

so

$$\mathrm{Var}(g) = a^2\sigma_{b_j}^2 + b^2\sigma_{a_j}^2 + \sigma_{a_j}^2\sigma_{b_j}^2 + 2ab\sigma_{a_j,b_j} + \sigma_{a_j,b_j}^2. \quad \text{(A5)}$$

Heterogeneity of the Total Effect Among Level 2 Units (Equation 8)

Kenny et al. (2003) represented variability in the total effect by

$$\mathrm{Var}(a_j b_j + c_j') = b^2\sigma_{a_j}^2 + a^2\sigma_{b_j}^2 + \sigma_{a_j}^2\sigma_{b_j}^2 + 2ab\sigma_{a_j,b_j} + \sigma_{a_j,b_j}^2 + \sigma_{c_j'}^2 + 2b\sigma_{a_j,c_j'} + 2a\sigma_{b_j,c_j'}. \quad \text{(A6)}$$

Again, in matrix notation,

$$\mathrm{Var}(g) = \mathbf{D}'\Sigma(\theta)\mathbf{D} + \tfrac{1}{2}tr\{(\mathbf{H}\Sigma(\theta))^2\}. \quad \text{(A7)}$$

When $g = a_j b_j + c_j'$, $(\delta g/\delta a)|_{a,b,c'} = b$, $(\delta g/\delta b)|_{a,b,c'} = a$, and $(\delta g/\delta c')|_{a,b,c'} = 1$, and hence $\mathbf{D}' = [b\ a\ 1]$. Here,

$$\Sigma(\theta) = \begin{bmatrix} \sigma_{a_j}^2 & \sigma_{a_j,b_j} & \sigma_{a_j,c_j'} \\ \sigma_{a_j,b_j} & \sigma_{b_j}^2 & \sigma_{b_j,c_j'} \\ \sigma_{a_j,c_j} & \sigma_{b_j,c_j'} & \sigma_{c_j'}^2 \end{bmatrix} \quad \text{(A8)}$$

and

$$\mathbf{H} = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad \text{(A9)}$$

so

$$\mathrm{Var}(g) = a^2\sigma_{b_j}^2 + b^2\sigma_{a_j}^2 + \sigma_{a_j}^2\sigma_{b_j}^2 + 2ab\sigma_{a_j,b_j} + \sigma_{a_j,b_j}^2 + \sigma_{c_j'}^2 + 2b\sigma_{a_j,c_j'}2a\sigma_{b_j,c_j'}. \quad \text{(A10)}$$

Sampling Variability of the Mean Estimated Indirect Effect (Equation 9)

The sampling variance of the mean estimated indirect effect may be obtained by similar procedures. The mean indirect effect is

$$g = ab + \sigma_{a_j b_j}. \quad \text{(A11)}$$

Using the second-order delta method requires the gradient and Hessian of g and the asymptotic covariance matrix of parameter estimates. Here, $\mathbf{D}' = [b\ a\ 1]$,

$$\mathbf{H} = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad \text{(A12)}$$

and

$$\Sigma(\theta) = \begin{bmatrix} \sigma^2_{\hat{a}} & \sigma_{\hat{a},\hat{b}} & 0 \\ \sigma_{\hat{a},\hat{b}} & \sigma^2_{\hat{b}} & 0 \\ 0 & 0 & \sigma^2_{\hat{\sigma}_{a_j,b_j}} \end{bmatrix} \quad \text{(A13)}$$

(asymptotic covariances between fixed and random effect parameters are zero). Thus,

$$\text{Var}(g) = \mathbf{D}'\Sigma(\theta)\mathbf{D} + \tfrac{1}{2}tr\{(\mathbf{H}\Sigma(\theta))^2\} = b^2\sigma^2_{\hat{a}} + a^2\sigma^2_{\hat{b}} + \sigma^2_{\hat{a}}\sigma^2_{\hat{b}} + 2ab\sigma_{\hat{a},\hat{b}} + (\sigma_{\hat{a},\hat{b}})^2 + \sigma^2_{\hat{\sigma}_{a_j,b_j}}. \quad \text{(A14)}$$

In practice, the sample estimates of these quantities are substituted for their population counterparts.

Sampling Variability of the Mean Estimated Total Effect (Equation 10)

The mean estimated total effect is

$$g = ab + c' + \sigma_{a_j,b_j}. \quad \text{(A15)}$$

Using the second-order delta method requires the gradient and Hessian of g and the asymptotic covariance matrix of parameter estimates. Here, $\mathbf{D}' = [b\ a\ 1\ 1]$,

$$\mathbf{H} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \quad \text{(A16)}$$

and

$$\Sigma(\theta) = \begin{bmatrix} \sigma^2_{\hat{a}} & \sigma_{\hat{a},\hat{b}} & \sigma_{\hat{a},\hat{c}'} & 0 \\ \sigma_{\hat{a},\hat{b}} & \sigma^2_{\hat{b}} & \sigma_{\hat{b},\hat{c}'} & 0 \\ \sigma_{\hat{a},\hat{c}'} & \sigma_{\hat{b},\hat{c}'} & \sigma^2_{\hat{c}'} & 0 \\ 0 & 0 & 0 & \sigma^2_{\hat{\sigma}_{a_j,b_j}} \end{bmatrix}. \quad \text{(A17)}$$

Thus,

$$\text{Var}(g) = \mathbf{D}'\Sigma(\theta)\mathbf{D} + \tfrac{1}{2}tr\{(\mathbf{H}\Sigma(\theta))^2\} = b^2\sigma^2_{\hat{a}} + a^2\sigma^2_{\hat{b}} + 2ab\sigma_{\hat{a},\hat{b}} + 2b\sigma_{\hat{a},\hat{c}'} + 2a\sigma_{\hat{b},\hat{c}'} + \sigma^2_{\hat{\sigma}_{a_j,b_j}} + \sigma^2_{\hat{c}'} + \sigma^2_{\hat{a}}\sigma^2_{\hat{b}} + (\sigma_{\hat{a},\hat{b}})^2. \quad \text{(A18)}$$

In practice, the sample estimates are substituted for their population counterparts.

Received March 16, 2005
Revision received March 1, 2006
Accepted March 1, 2006 ■