Psychological Methods
2007, Vol. 12, No. 1, 23–44

1082-989X/07/$12.00 DOI: 10.1037/1082-989X.12.1.23

Bias in Cross-Sectional Analyses of Longitudinal Mediation

Scott E. Maxwell
University of Notre Dame

David A. Cole
Vanderbilt University

Most empirical tests of mediation utilize cross-sectional data despite the fact that mediation consists of causal processes that unfold over time. The authors considered the possibility that longitudinal mediation might occur under either of two different models of change: (a) an autoregressive model or (b) a random effects model. For both models, the authors demonstrated that cross-sectional approaches to mediation typically generate substantially biased estimates of longitudinal parameters even under the ideal conditions when mediation is complete. In longitudinal models where variable *M* completely mediates the effect of *X* on *Y*, cross-sectional estimates of the direct effect of *X* on *Y*, the indirect effect of *X* on *Y* through *M*, and the proportion of the total effect mediated by *M* are often highly misleading.

Keywords: mediation, direct effect, indirect effect, cross-sectional designs, longitudinal designs

Mediation is of fundamental interest in many areas of psychology because of the central role it can play in answering questions about underlying processes. After discovering a relation between two variables, investigators frequently seek to clarify the mechanism (or mediational process) underlying this relation. Over the past 20 years, most efforts to test for mediation have been based on cross-sectional data and have involved methods described by Baron and Kenny (1986) and elaborated by numerous authors (e.g., Kenny, Kashy, & Bolger, 1998; MacKinnon, Lockwood, Hoffman, West, & Sheets, 2002; Shadish, Cook, & Campbell, 2002; Shrout & Bolger, 2002). Although most methodological discussions and substantive investigations of mediation have ignored any consideration of time sequence, a few recent articles have begun to consider the explicit role of time in studying mediational processes (Cole & Maxwell, 2003; Collins, Graham, & Flaherty, 1998; Kenny, Korchmaros, & Bolger, 2003; MacCallum & Austin, 2000; Tein, Sandler, MacKinnon, & Wolchik, 2004). An earlier but very important strand of research was established by Gollob and Reichardt (1985, 1987, 1991), who emphasized the importance of time in the formation and interpretation of structural equation models. Despite these calls for longitudinal approaches to mediation, the modal methodology for testing mediation continues to be cross-sectional. The overarching goal of this article is to describe the circumstances under which and the degree to which cross-sectional efforts to estimate mediation will be biased and potentially seriously misleading.

How prevalent is the use of cross-sectional designs to test mediation? To address this question, we reviewed the literature. Using PsycINFO, we discovered that the five American Psychological Association journals most likely to have published articles that tested mediational hypotheses were the *Journal of Personality and Social Psychology*, the *Journal of Consulting and Clinical Psychology*, the *Journal of Applied Psychology*, *Health Psychology*, and *Developmental Psychology*. In 2005 alone, these journals collectively published 68 articles (containing 72 studies) that described tests of mediation in their titles or abstracts. Of these, 28 (39%) were based on completely cross-sectional designs. Another 10 (14%) ignored or abused the longitudinal structure of their own data by focusing on only a single wave, averaging across waves, or treating later variables as predictors of earlier variables when testing for mediation. Taken together, 53% of the reviewed studies were essentially cross-sectional in that they tested mediation with methods that did not allow time for an independent variable to have an effect on a dependent variable. Within the remaining truly longitudinal studies, other problems arose. For example, 27 (38% of all mediational studies) were what

Editor's Note. William Shadish served as the action editor for this article.—SGW

Scott E. Maxwell, Department of Psychology, University of Notre Dame; David A. Cole, Department of Psychology and Human Development, Vanderbilt University.

This research was supported in part by National Institute of Mental Health Grant R01MH64650 and National Institute of Child Health and Human Development Grant P30HD15052 to Vanderbilt University. We thank Ken Kelley and Joseph Rausch for comments on a draft of this article.

Correspondence concerning this article should be addressed to Scott E. Maxwell, Department of Psychology, University of Notre Dame, Notre Dame, IN 46556. E-mail: smaxwell@nd.edu

we have called *half-longitudinal designs* (Cole & Maxwell, 2003), insofar as time elapsed either between the measurement of X (the presumed causal variable) and M (the presumed mediator) or between M and Y (the presumed dependent variable), but not both. In other words, two of the three variables were obtained concurrently. Among the remaining 7 studies that had fully longitudinal designs, only one (a) controlled for prior levels of M when testing the association between X at Time 1 and the mediator at Time 2 and (b) controlled for prior levels of Y when testing the association between the mediator at Time 2 and the dependent variable at Time 3. This review reveals two important things: (a) that cross-sectional tests of mediation are still the norm in premier journals from a diversity of psychological disciplines and (b) that mediational tests of longitudinal data only rarely reflect recent methodological advances (e.g., Cole & Maxwell, 2003; Collins et al., 1998; Kenny et al., 2003; Tein et al., 2004; see also Gollob & Reichardt, 1991).

Why have cross-sectional analyses of mediational processes persisted despite methodological arguments that longitudinal designs are more appropriate? One factor is that little is known about the practical consequences of using cross-sectional designs to study mediation. If differences between longitudinal and cross-sectional analyses of mediation are small in practice, researchers might be justified in continuing to study mediation with cross-sectional designs despite a theoretical disadvantage. On the other hand, if differences are likely to be large in practice, longitudinal designs may be necessary.

This article has two goals. The first is to establish conditions under which cross-sectional analyses of mediation can be expected to yield the same conclusion as longitudinal designs. In particular, we derive specific conditions under which cross-sectional analyses can be trusted to provide accurate inferences about longitudinal mediational processes. If substantive researchers can justify these specific conditions, longitudinal designs are unnecessary. However, in many situations, these conditions will be difficult to justify, in which case cross-sectional analyses will yield different results from longitudinal analyses. This leads to our second goal, which is to examine the magnitude of this difference. If the difference is small, cross-sectional analyses might be justified as providing a good approximation to a longitudinal analysis. However, if the difference is large, any interpretation of cross-sectional analyses may be questionable. Thus, a major goal of this article is to ascertain the extent to which cross-sectional designs are likely to provide a good approximation to underlying longitudinal mediational processes.

Two Models of Change

By its very definition, mediation implies change over time. Some variable X influences a mediator M, which in turn influences an outcome variable Y. In this article, we focus on mediational processes where all three variables, X, M, and Y, potentially change over time. For example, parental depression (X) might influence parenting behavior (M), which then promotes child depression (Y). Furthermore, we expect parental depression as well as parenting behavior and child depression to change over time.[1] Although change is central to the concept of mediation, it is nevertheless true in virtually all behavioral studies that any given variable measured at time t correlates with itself when measured at a later time $t + 1$. We use the term *stability* to represent this property, recognizing that this definition implies some degree of rank-order correlation over time but does not directly imply any restrictions on means or standard deviations. For example, depression scores measured in April are likely to correlate with depression scores measured 6 months later in October. If this correlation were large, one would say that depression is stable even if the mean depression score changes during this period. Thus, any viable model of change must allow variables to demonstrate some stability over time. Lord (1967, 1969) pointed out nearly 40 years ago that there are two fundamentally different conceptualizations of change, both of which allow variables to exhibit stability over time.

One model of change stipulates that the value of a variable at some future time $t + 1$ depends at least in part on the value of that same variable at some earlier time t. This type of conceptualization typically is translated into an autoregressive statistical model of change. In the simplest case, the value of Y for any given individual i at time $t + 1$ is a linear function of that same person's score on Y at time t:

$$Y_{it+1} = \beta_0 + \beta_1 Y_{it} + \varepsilon_{it}. \tag{1}$$

As long as β_1 is nonzero, Y necessarily exhibits at least some degree of stability over time. By allowing stability to be represented as a parameter in the model, it is unnecessary to stipulate in advance the precise correlation between Y at time t and Y at time $t + 1$. In many situations, a central question of interest is the extent to which future values of Y depend not only on prior values of Y but also on other variables. For example, introducing an additional variable M into the model shown in Equation 1 provides an estimate of the extent to which M is related to future values of Y, controlling for earlier values of Y:

$$Y_{it+1} = \beta_0 + \beta_1 Y_{it} + \beta_2 M_{it} + \varepsilon_{it}. \tag{2}$$

[1] A different situation arises where only M and Y would be expected to change over time. For example, X might represent group membership, which could be stable over time. Groups might either be naturally occurring, such as sex, or formed by the experimenter, ideally through random assignment. In either case, it is important to note that this is a different design than we consider here.

The key point here is that the influence of M on changes in Y can be assessed by forming a regression model in which both M and Y at time t are used as predictors of Y at time $t + 1$. This basic idea can easily be extended to latent-variable models, where multiple indicators are available for M at time t and for Y at times t and $t + 1$.

A second model of change stipulates that each person has some potentially unique trajectory of change over time. This type of model goes by various names, including random effects, multilevel, and latent growth curve models. We use the term *random effects model* with the understanding that in most situations, these various terms are essentially interchangeable. From the random effects perspective, an individual's score on some variable Y can be expressed as a function of time. In the simple case of straight-line growth, a suitable model can be written as

$$Y_{it} = \beta_{0i} + \beta_{1i} t_i Y_{it} + \varepsilon_{it}. \tag{3}$$

Although Equation 3 bears some resemblance to Equation 1, there are two critical differences. First, in the autoregressive model of Equation 1, Y at any time point depends at least partly on Y at an earlier time point. In contrast, in the random effects model of Equation 3, Y does not depend directly on previous values of Y but instead depends simply on the time of measurement. At first glance, this new formulation might seem incapable of allowing Y to exhibit the stability over time that variables almost always display. However, the second critical difference between this formulation and the autoregressive model allows for stability. Notice that each model parameter (i.e., β_{0i} and β_{1i}) in Equation 3 contains an i subscript, unlike the parameters of the model in Equation 1. The presence of this subscript reflects the fact that each individual is allowed to have his or her own intercept and slope. As a consequence, each person potentially has a unique trajectory of change over time. In other words, individuals vary in their initial values of Y and also vary in their rate of change in Y over time. Such variability in these random effects β_{0i} and β_{1i} allows scores on variable Y at time $t + 1$ to correlate with scores on Y at time t. Thus, the random effects model, like the autoregressive model, is capable of representing data where variables exhibit stability over time.

Each of these two models of change has its proponents. Although we describe situations where each type of model might be particularly appropriate, we do not attempt to adjudicate which model is better. Instead, we consider each model in turn. First, we presume that the autoregressive model is the appropriate model of change for certain phenomena. From this perspective, we consider a model of mediation. Our key question is the extent to which cross-sectional designs can be relied upon to provide an accurate reflection of mediation if the true process follows an autoregressive model. Second, we presume that the random effects model is a more realistic model of change for other types of phenomena. We then consider mediation from this perspective. Our key question is the extent to which cross-sectional designs can be relied upon to provide an accurate reflection of mediation when the true process follows a random effects model.

Mediation From the Perspective of an Autoregressive Model of Change

We begin our consideration of the autoregressive model with a hypothetical example. A developmental psychopathologist might wonder why parental depression is associated with child depression. One possibility is that depressed parents engage in problematic parenting practices, which in turn foster depression in their children. Implicit in this answer is the view that a third variable (problematic parenting) at least partially explains the association between two other variables (parental depression and child depression). The simplest design for studying mediation involves obtaining a measure of each of the three variables. A typical statistical model of the hypothesized relations among these variables is depicted in Figure 1. This model suggests that parental depression (X) has some influence on problematic parenting (M), and both parental depression and problematic parenting may have some direct effect on childhood depression (Y). The direct effect of X on Y is reflected by c'. Of course, X may also have an indirect effect on Y through M, represented by the product $a'b'$ in the model. In the current example, the key research question is whether problematic parenting behavior mediates the effect of parental depression on child depression.

In any such test, one must consider the advantages of obtaining multiple measures of each construct and using structural equation modeling. Such an approach enables the investigator to avoid the bias that is typically produced by random error of measurement. Throughout the current article, we assume that the variables are latent or are otherwise measured without error. In other words, we assume a best

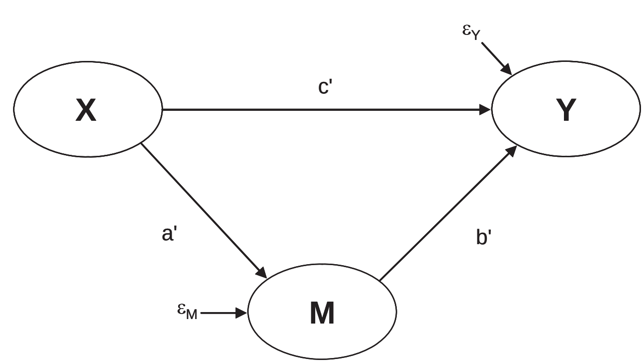

Figure 1. Cross-sectional mediation model. X = independent variable; M = mediator; Y = dependent variable.

case scenario of variables measured with no error either because they are latent variables or because they reflect certain manifest variables such as gender and age that can be measured perfectly.[2]

To address this question (i.e., whether problematic parenting mediates the effect of parental depression on child depression), let us imagine that a researcher first obtains cross-sectional data from 100 families on all three variables. Table 1 contains correlations between these three variables that might derive from such a study. The results of a standard mediational analysis using ordinary least squares regression are shown in Figure 2. Several conclusions can be drawn here: (a) The independent variable does significantly predict the mediator, (b) the mediator does not significantly predict the dependent variable, and (c) the independent variable remains a significant predictor of the dependent variable even when controlling for the mediator. From these results, we would infer that problematic parenting is not a mediator (or, more carefully stated, we would infer that we have no clear basis for concluding that problematic parenting is a mediator of the relation between parental depression and child depression).

Figure 2 shows that the estimated direct effect of parental depression on child depression is 0.22. The estimated indirect effect is the product of 0.37 and 0.15, which equals 0.05. The estimated total effect, which is the sum of the direct effect and the indirect effect, equals 0.27. The proportion of the total effect mediated by problematic parenting is symbolized as P_M and computed as the ratio of the indirect effect (0.05) to the total effect (0.27); here, P_M equals 0.19. Thus, the single best estimate from the observed data is that 19% of the total effect of parental depression on child depression can be attributed to problematic parenting, whereas the remaining 81% cannot be explained by problematic parenting.[3]

This example reflects a cross-sectional design in which all three variables have been measured simultaneously, probably the most common type of mediational study in psychology. In reality, however, causal effects such as those depicted in Figures 1 and 2 obviously occur over time. That is, some amount of time must elapse between the cause and its effect (Cohen, Cohen, West, & Aiken, 2003; Collins et al., 1998; Gollob & Reichardt, 1985, 1987, 1991). A depiction

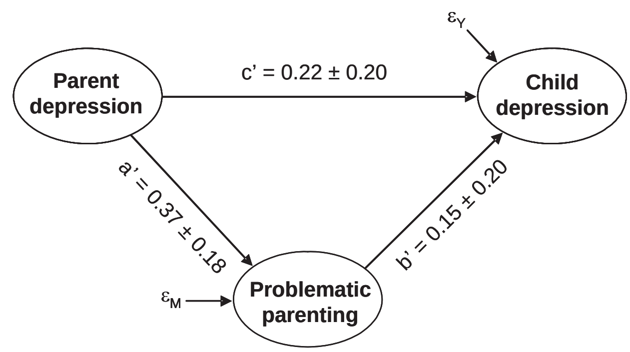

Figure 2. Cross-sectional model parameter estimates (and 95% confidence intervals) based on the correlation matrix from Table 1.

of how X, M, and Y might relate to one another over time is shown in Figure 3, which is an example of an autoregressive model. Unlike the models shown in Figures 1 and 2, the longitudinal model in Figure 3 explicitly depicts the time sequence relating the cause (X) to the mediator (M) and the effect (Y). In this path diagram, the autoregressive effects of X, M, and Y are represented by paths x, m, and y, respectively. The direct effects of $X \rightarrow M$, $M \rightarrow Y$, and $X \rightarrow Y$ are represented by paths a, b, and c, respectively.

Continuing our hypothetical example, let us imagine that the investigator is undaunted by the previous null results. Indeed, compelled by the need for longitudinal data, the researcher extends the cross-sectional study for two additional waves, obtains data on the same 100 individuals (and on the same three variables), and turns the previous cross-sectional study into a three-wave panel design. The correlations for this study appear in Table 2. The cross-sectional

Table 1
Correlation Matrix for Hypothetical Data Relating Parental Depression, Problematic Parenting Behavior, and Child Depression

Variable	1	2	3
1. Parental depression (X)	1.00		
2. Problematic parenting (M)	.37	1.00	
3. Child depression (Y)	.27	.23	1.00

Note. X = independent variable; M = mediator; Y = dependent variable.

[2] Our purpose in this article is to evaluate the performance of cross-sectional measures of mediation under ideal conditions. To the extent that fallible manifest variables form the basis of cross-sectional analyses, known biases will occur in estimating relevant model parameters. Of course, it is possible that multiple sources of bias could counteract one another, but such wishful thinking is rarely correct. In almost all circumstances, the types of problems with cross-sectional efforts to test longitudinal mediation generalize to situations in which X, M, and Y are measured with error.

[3] Shrout and Bolger (2002) and MacKinnon, Lockwood, and Williams (2004) have shown that a bias-corrected bootstrap confidence interval for the indirect effect can be more informative than simply testing the statistical significance of each path in the presumed structural model. Forming such a 95% confidence interval for these data reveals that the indirect effect of parental depression on child depression through problematic parenting (i.e., the effect of X on Y through M) plausibly ranges from −0.01 to 0.15. Thus, it is plausible that the population indirect effect equals zero, and we have no clear basis for concluding that problematic parenting mediates the relation between parental depression and child depression.

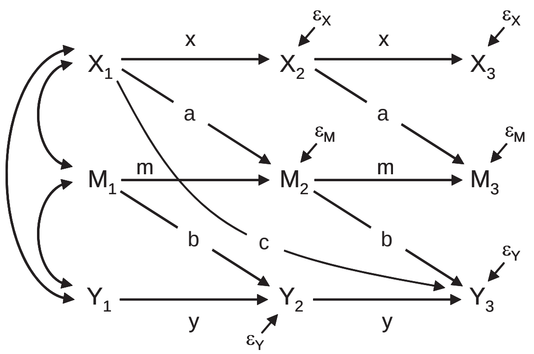

Figure 3. Longitudinal mediation model. X = independent variable; M = mediator; Y = dependent variable.

correlations for Waves 2 and 3 are essentially the same as they were for Wave 1. Consequently, the cross-sectional mediational analyses at Wave 2 and at Wave 3 yield essentially the same disappointing results as those at Wave 1. Structural equation analysis reveals that the longitudinal model in Figure 3 fits the data well, $\chi^2(22, N = 100) = 23.06$, $p = .40$, root-mean-square error of approximation = .022, and generates the results shown in Figure 4. These longitudinal results yield several important conclusions: (a) The independent variable (parental depression) significantly predicts the mediator (problematic parenting), (b) the mediator (problematic parenting) significantly predicts the dependent variable (child depression), and (c) the direct effect of parental depression on child depression becomes nonsignificant after controlling for problematic parenting, $t = 0.17$, $p = .86$. The standardized regression coefficient from parental depression to child depression is only 0.01. The indirect effect *ab* is 0.08, which is 88% of the total effect of 0.09. In other words, an estimated 88% of the total effect is mediated by problematic parenting. By all accounts, these longitudinal conclusions completely contradict the cross-sectional conclusions—even when the cross-sectional approach is applied to data derived from the larger longitudinal study.[4] Clearly, the cross-sectional results can be highly misleading. Even more disturbing, however, is the fact that the problem is not limited to this example but extends to almost all cross-sectional efforts to estimate mediational parameters.

This example highlights the need to understand more completely the implications of taking a cross-sectional approach to what is essentially a longitudinal process. In the remainder of this article, we examine the implications of longitudinal processes on (a) estimates of the cross-sectional direct effect of X on Y, (b) estimates of the cross-sectional indirect effect of X on Y through M, and (c) cross-sectional estimates of proportion of total effect mediated by M. We do so first by considering each of these estimates from the perspective of an underlying autoregressive longitudinal model. We then consider the same three estimates from the perspective of an underlying random effects longitudinal model. In all cases, we rely on mathematical derivations of possible biases that can result from cross-sectional analyses of true longitudinal processes. As we explain later, such bias immediately renders any statistical tests or confidence intervals of questionable value. If cross-sectional analyses are shooting at the wrong target, statistical tests and confidence intervals will necessarily be distorted. Thus, our focus throughout is twofold: (a) ascertaining when cross-sectional parameters differ from parameters in longitudinal models and (b) quantifying the magnitude of these differences.

Estimating the Cross-Sectional Direct Effect of X on Y: Autoregressive Model

When might it seem reasonable to assess mediational processes using cross-sectional data? Let us consider an expanded version of the longitudinal example examined above. Figure 5 depicts such a model in which the mediational processes extend over t waves. As in our cross-sectional example, we focus on the specific case in which there is no direct path from X to Y (i.e., path $c = 0$), indicating that M completely mediates the $X \rightarrow Y$ relation at every time interval. In many situations, dynamic relations such as these may have stabilized; the path coefficients connecting any pair of variables at adjacent time points are the same regardless of the specific time period. In addition, the system itself may have reached equilibrium such that cross-sectional correlations among X, M, and Y are the same at every time point. To the extent that both of these assumptions hold, a longitudinal analysis may seem unnecessary because the only reason cross-sectional correlations among X, M, and Y would differ from one time point to another would be because of sampling error. Thus, the only apparent value in obtaining measures at multiple time points might seem to be the fact that one would derive more precise estimates due to simple replication over time. Under such conditions, a cross-sectional mediational analysis might seem entirely reasonable; the benefits gained from a longitudinal study might not seem to justify the extra cost and effort. However, we show below that even when these conditions are met, cross-sectional analyses very rarely re-

[4] An expanded version of this longitudinal model might include a path that allows X to cause Y directly at a time lag of 1 time unit, instead of 2 units. This model also provides a good fit to the data. However, both direct paths from parental depression to child depression are essentially zero in this model. Furthermore, a model with no direct paths whatsoever from parental depression to child depression also fits these data very well. Thus, all the results of all three models agree that it is plausible that parental depression has no direct effect on child depression.

Table 2
Correlation Matrix for Hypothetical Longitudinal Data Relating Parental Depression (X), Problematic Parenting (M), and Child Depression (Y) Across Three Waves

Variable	1	2	3	4	5	6	7	8	9
1. *X*1	1.00								
2. *M*1	.37	1.00							
3. *Y*1	.27	.23	1.00						
4. *X*2	.89	.30	.26	1.00					
5. *M*2	.39	.42	.12	.36	1.00				
6. *Y*2	.29	.48	.76	.28	.21	1.00			
7. *X*3	.79	.31	.21	.88	.27	.22	1.00		
8. *M*3	.33	.26	.08	.43	.44	.15	.37	1.00	
9. *Y*3	.33	.39	.66	.31	.44	.78	.29	.22	1.00

Note. The three uppermost correlations are the same as those in Table 1. X = independent variable; M = mediator; Y = dependent variable.

veal the true nature of mediational processes as conceptualized in the longitudinal model of Figure 5.

Appendix A shows the derivations of the cross-sectional zero-order correlations among X, M, and Y from the model shown in Figure 5. These derivations assume that (a) M at time t fully mediates the relation between X at time $t - 1$ and Y at time $t + 1$, (b) the path coefficients a, b, x, m, and y are invariant over time, and (c) the system has reached equilibrium so that the cross-sectional correlations among X, M, and Y do not depend on the time of measurement. Under these conditions, Appendix A shows that the population cross-sectional correlations are given by

$$\rho_{X_tM_t} = \frac{ax}{1 - mx}, \tag{4}$$

$$\rho_{X_t,Y_t} = \frac{abx^2}{(1 - mx)(1 - xy)}, \text{ and} \tag{5}$$

$$\rho_{M_tY_t} = \frac{bm}{(1 - my)} + \frac{a^2bx}{(1 - mx)(1 - my)(1 - xy)}. \tag{6}$$

The essential question is how well do the cross-sectional parameters a', b', and c' as shown in Figure 1 accurately represent the underlying longitudinal mediational process. We begin with c', which represents the direct effect of X on Y controlling for M. The population value of c' in the cross-sectional analysis can be derived from the correlations shown in Equations 4 through 6:

$$c' = \frac{\rho_{X_tY_t} - \rho_{X_tM_t}\,\rho_{M_tY_t}}{1 - \rho^2_{X_tM_t}}, \tag{7}$$

$$c' = \frac{\dfrac{abx^2}{(1 - mx)(1 - xy)} - \left\{\dfrac{ax}{(1 - mx)}\left[\dfrac{bm}{(1 - my)} + \dfrac{a^2bx}{(1 - mx)(1 - my)(1 - xy)}\right]\right\}}{1 - \rho^2_{X_tM_t}}, \text{ and} \tag{8}$$

$$c' = \frac{abx[\rho_{X_tX_{t-1}} - \rho_{M_tM_{t-1}}]}{(1 - mx)(1 - my)(1 - xy)(1 - \rho^2_{X_tM_t})}, \tag{9}$$

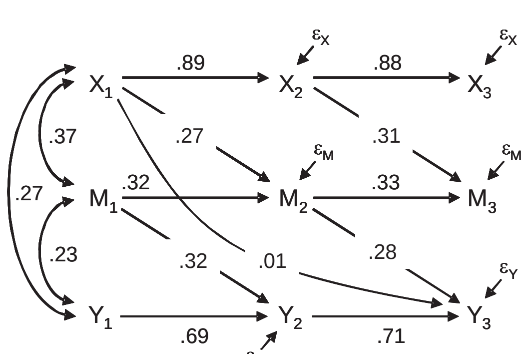

Figure 4. Longitudinal parameter estimates based on Table 2. X = independent variable; M = mediator; Y = dependent variable.

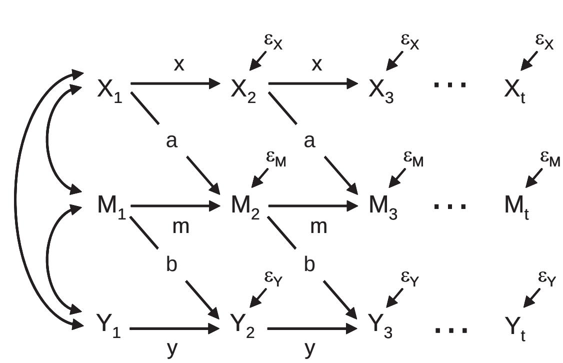

Figure 5. Extended longitudinal model of complete mediation, where path c (the direct effect of X on Y) is zero. X = independent variable; M = mediator; Y = dependent variable.

where $\rho_{X_tX_{t-1}}$ and $\rho_{M_tM_{t-1}}$ refer to the stability of X and M, respectively, between two adjacent time points. In other words, $\rho_{X_tX_{t-1}}$ is the correlation between X at any time t and X at the previous time point $t - 1$. Similarly, $\rho_{M_tM_{t-1}}$ is the correlation between M at any time t and M at the previous time point $t - 1$.

Assuming that $a > 0$, $b > 0$, and $x > 0$, it follows from Equation 9 that

$$c' = 0 \text{ if and only if } \rho_{X_tX_{t-1}} = \rho_{M_tM_{t-1}}, \quad (10)$$

$$c' > 0 \text{ if and only if } \rho_{X_tX_{t-1}} > \rho_{M_tM_{t-1}}, \text{ and} \quad (11)$$

$$c' < = 0 \text{ if and only if } \rho_{X_tX_{t-1}} < \rho_{M_tM_{t-1}}. \quad (12)$$

The import of these expressions becomes strikingly evident when one notes that c' represents the direct effect of X on Y in the cross-sectional analysis and that the underlying causal model shown in Figure 5 has no direct effect of X on Y whatsoever. Thus, the parameter value c' in the cross-sectional analysis is correct only when it equals zero, which is true only when X and M are equally stable. In other words, researchers who analyze cross-sectional mediational data are estimating an appropriate direct effect of X on Y (from a longitudinal perspective) if and only if X and M are equally stable. Recall that this means that the population correlation between two measures of X at adjacent time periods must equal the population correlation between two measures of M at adjacent time periods. In our experience, this assumption is often dubious and rarely (if ever) tested. When this condition is not met, cross-sectional analyses will generate biased estimates of the true direct effect of X on Y no matter how large a sample size is used.

Although bias is generally undesirable in any statistical procedure, a saving grace in some situations is that the direction of bias can be known with certainty. For example, random error of measurement necessarily lowers the absolute value of the correlation between two variables, so the population correlation between manifest variables must be closer to zero than the population correlation between corresponding latent variables. Unfortunately, in the current mediation example, Expressions 11 and 12 show that the direction of bias that derives from using cross-sectional data to estimate direct longitudinal effects is unknowable without more information about the relative stability of X and M. If X is more stable than M, the direct effect will be positive when it should be zero. However, if X is less stable than M, the direct effect will be negative when it should be zero.

Our example of parental depression, problematic parenting, and child depression illustrates how bias can arise from a cross-sectional analysis. Recall that the cross-sectional analysis implied that 19% of the total effect of parental depression on child depression was mediated by problematic parenting; however, the corresponding longitudinal analysis revealed that problematic parenting actually mediated 100% of the effect of parental depression on child depression. In this example, the cross-sectional analysis is based upon a subset of the same correlations that gave rise to the longitudinal results (compare Tables 1 and 2). Both correlation matrices are completely consistent with the longitudinal model of Figure 5 with parameter values of $x = 0.90$, $a = 0.30$, $b = 0.30$, $y = 0.70$, and $m = 0.30$. In the model that generated these data, path $c = 0$, and mediation was 100% complete. In stark contrast, the cross-sectional approach suggests that mediation was only 19% complete.[5]

Why have we been misled by the cross-sectional analysis? The reason is that in this case, X (parental depression) has a stability coefficient of 0.90, whereas the mediator has a stability coefficient of 0.41. Expression 11 shows that when $\rho_{X_tX_{t-1}} > \rho_{M_tM_{t-1}}$, $c' > 0$ under complete longitudinal mediation. Indeed, Figure 2 shows that $c' = 0.22$ for these data even though there is no direct effect of parental depression on child outcome in the longitudinal model.

Expressions 10, 11, and 12 establish that cross-sectional analyses can easily generate biased estimates of longitudinal direct effects. This problem might be negligible if the magnitude of bias were small. Tables 3 and 4 provide population values for c' (and c) under a variety of plausible conditions. The rightmost column in Tables 3 and 4 shows that the bias can be quite substantial. These bias values are easily interpreted as they represent the difference between the population standardized regression coefficient (c') in a cross-sectional design and the true value of this effect ($c = 0$) in the underlying longitudinal model.

Table 3 depicts situations where X is at least as stable as M. As implied by Expressions 10 and 11, values of the cross-sectional parameter c' are necessarily nonnegative in these cases. Even though there is no direct effect of X on Y in the longitudinal model, the corresponding direct effect in the cross-sectional model is almost always positive. Only in the (unlikely) special case where X and M are equally stable (both equal to .70 in Table 3) is the cross-sectional direct

[5] The statement that problematic parenting accounts for 100% of the effect of parental depression on child depression may seem to contradict our earlier statement that the single best estimate in our numerical example is that problematic parenting accounts for 88% of the effect of parental depression on child depression. The 100% figure is the true population value, whereas the 88% figure is the value obtained in this specific sample. Thus, the discrepancy between these two values simply reflects sampling error. We could have created data where, even in the sample, problematic parenting accounted for exactly 100% of the effect of parental depression on child depression, but we decided that it would be less misleading to allow for a certain amount of sampling error in our numerical example instead of having sample values literally duplicate population values.

Table 3
Bias in the Estimated Direct Effect of X on Y When X Is at Least as Stable as M and the True Direct Effect (Longitudinal Path c) Equals Zero

Stability		Longitudinal parameters[a]						Cross-sectional parameters[b]				Bias
$\rho_{X_tX_{t-1}}$	$\rho_{M_tM_{t-1}}$	a	b	c	x	m	y	ρ_{XM}	ρ_{MY}	ρ_{XY}	c'	$(c' - c)$
1.0	0.7	0.5	0.4	0.0	1.0	0.33	0.5	0.74	0.51	0.60	0.48	0.48
1.0	0.5	0.5	0.4	0.0	1.0	0.19	0.5	0.62	0.36	0.49	0.44	0.44
1.0	0.3	0.5	0.4	0.0	1.0	0.04	0.5	0.52	0.23	0.42	0.41	0.41
0.9	0.7	0.5	0.4	0.0	0.9	0.36	0.5	0.67	0.48	0.44	0.22	0.22
0.9	0.5	0.5	0.4	0.0	0.9	0.22	0.5	0.56	0.33	0.37	0.27	0.27
0.9	0.3	0.5	0.4	0.0	0.9	0.06	0.5	0.48	0.20	0.31	0.28	0.28
0.8	0.7	0.5	0.4	0.0	0.8	0.40	0.5	0.59	0.45	0.32	0.08	0.08
0.8	0.5	0.5	0.4	0.0	0.8	0.25	0.5	0.50	0.30	0.27	0.15	0.15
0.8	0.3	0.5	0.4	0.0	0.8	0.09	0.5	0.43	0.19	0.23	0.18	0.18
0.7	0.7	0.5	0.4	0.0	0.7	0.45	0.5	0.51	0.43	0.22	0.00	0.00
0.7	0.5	0.5	0.4	0.0	0.7	0.28	0.5	0.44	0.29	0.19	0.08	0.08
0.7	0.3	0.5	0.4	0.0	0.7	0.11	0.5	0.38	0.17	0.16	0.12	0.12

Note. X = independent variable; M = mediator; Y = dependent variable; a = direct effect of $X \rightarrow M$; b = direct effect of $M \rightarrow Y$; c = direct effect of $X \rightarrow Y$; x = autoregressive effect of X; m = autoregressive effect of M; y = autoregressive effect of Y; c' = cross-sectional direct effect of $X \rightarrow Y$.
[a] Hypothetical path coefficients for the longitudinal model depicted in Figure 5. [b] The cross-sectional parameters that would emerge for the model in Figure 1 if the longitudinal model in Figure 5 were the true model.

effect parameter c' equal to zero. Most importantly from a practical perspective, the value of c' can be quite substantial even in the population. Thus, even in very large samples, a cross-sectional analysis can yield evidence of a medium to large direct effect of X on Y even when the actual direct effect in the longitudinal model is zero.

Table 4 depicts situations where M is at least as stable as X. As implied by Expression 12, values of the cross-sectional parameter c' are necessarily negative whenever M is more stable than X. Even though there is no direct effect of X on Y in the longitudinal model, the corresponding direct effect parameter in the cross-sectional model is negative. Only in the special case where X and M are equally stable (both equal to .70 in Table 4) is the cross-sectional direct effect parameter c' equal to zero. All other cases generate negatively biased values for the direct effect of X on Y. The magnitude of bias shown in Table 4 is generally less than that in Table 3. Even so, some of the negative values in Table 4 are large enough (in absolute value) to cause consternation in a researcher who would probably be perplexed by an analysis suggesting a negative direct effect among variables all of which are positively related. Our analysis shows that such an occurrence may simply reflect the kind of bias that can arise from the cross-sectional analysis of a longitudinal phenomenon. Except for the special (and unusual) case where the independent variable X and the me-

Table 4
Bias in the Estimated Direct Effect of X on Y When M Is at Least as Stable as X and the True Direct Effect (Longitudinal Path c) Equals Zero

Stability		Longitudinal parameters[a]						Cross-sectional parameters[b]				Bias
$\rho_{X_tX_{t-1}}$	$\rho_{M_tM_{t-1}}$	a	b	c	x	m	y	ρ_{XM}	ρ_{MY}	ρ_{XY}	c'	$(c' - c)$
0.7	0.9	0.5	0.4	0.0	0.7	0.60	0.5	0.60	0.61	0.26	−0.17	−0.17
0.7	0.8	0.5	0.4	0.0	0.7	0.52	0.5	0.55	0.51	0.24	−0.07	−0.07
0.7	0.7	0.5	0.4	0.0	0.7	0.45	0.5	0.51	0.43	0.22	0.00	0.00
0.6	0.9	0.5	0.4	0.0	0.6	0.65	0.5	0.49	0.60	0.17	−0.17	−0.17
0.6	0.8	0.5	0.4	0.0	0.6	0.57	0.5	0.46	0.50	0.16	−0.09	−0.09
0.6	0.7	0.5	0.4	0.0	0.6	0.49	0.5	0.42	0.42	0.15	−0.04	−0.04
0.5	0.9	0.5	0.4	0.0	0.5	0.71	0.5	0.39	0.60	0.10	−0.15	−0.15
0.5	0.8	0.5	0.4	0.0	0.5	0.62	0.5	0.36	0.50	0.10	−0.10	−0.10
0.5	0.7	0.5	0.4	0.0	0.5	0.53	0.5	0.34	0.41	0.09	−0.06	−0.06

Note. X = independent variable; M = mediator; Y = dependent variable; a = direct effect of $X \rightarrow M$; b = direct effect of $M \rightarrow Y$; c = direct effect of $X \rightarrow Y$; x = autoregressive effect of X; m = autoregressive effect of M; y = autoregressive effect of Y; c' = cross-sectional direct effect of $X \rightarrow Y$.
[a] Hypothetical path coefficients for the longitudinal model depicted in Figure 5. [b] The cross-sectional parameters that would emerge for the model depicted in Figure 1 if the longitudinal model in Figure 5 were the true model.

diator M are equally stable, analysis of cross-sectional correlations will tend to indicate the presence of a direct effect of X on the dependent variable Y even when no such direct effect exists from a longitudinal perspective.

Estimating the Cross-Sectional Indirect Effect of X on Y Through M: Autoregressive Model

The previous sections have demonstrated the potential for serious bias in cross-sectional estimation of the direct effect of X on Y. We now turn our attention to cross-sectional estimation of the indirect effect. Previously, we have shown that a necessary and sufficient condition for the cross-sectional analysis to correctly identify the absence of a longitudinal direct effect of X on Y is that X and M be equally stable. However, even this condition does not guarantee that the cross-sectional analysis will accurately reflect the magnitude of the longitudinal indirect effect of X on Y through M. In fact, the indirect effect $a'b'$ in the cross-sectional model generally provides a poor estimate of the longitudinal indirect effect ab even under the best of circumstances.

Appendix B shows that when longitudinal path c equals zero, the difference between the indirect effect ab in the longitudinal model and $a'b'$ in the cross-sectional analysis can be written as

$$a'b' - ab = \frac{ab[x^2 - (1 - mx)(1 - xy)]}{(1 - mx)(1 - xy)} - c'. \qquad (13)$$

Even if X and M are equally stable (in which case, the cross-sectional path c' also equals zero), $a'b'$ will equal ab if and only if one of three additional conditions holds: (a) $a = 0$, (b) $b = 0$, or (c) $x^2 = (1 - mx)(1 - xy)$. Mediation is not possible if either a or b equals 0, and the third condition appears to have no straightforward interpretation or meaning. Thus, even if X and M are equally stable and even if there is complete mediation, the cross-sectional analysis will almost never reflect the longitudinal indirect effect accurately.

The panels of Figure 6 graph the difference between the cross-sectional indirect effect $a'b'$ and the longitudinal indirect effect ab for selected parameter values. These graphs show that even when X and M are equally stable, the cross-sectional indirect effect generally does not equal the indirect effect of the longitudinal model. As implied by Equation 13, the graphs also show that the direction of the bias can be either positive or negative. Furthermore, the graphs reveal that the magnitude of bias can be substantial in either direction. When X and M exhibit low stability (i.e., when $\rho_{X_tX_{t-1}}$ and $\rho_{M_tM_{t-1}}$ are close to .1), the cross-sectional indirect effect can be as much as 0.30 to 0.40 less than the longitudinal indirect effect. Conversely, when X and M are relatively stable (i.e., when $\rho_{X_tX_{t-1}}$ and $\rho_{M_tM_{t-1}}$ are close to .9), the cross-sectional indirect effect can be 0.40 to 0.50 more than the longitudinal indirect effect. We hasten to note that the values shown in Figure 6 reflect situations where X and M are equally stable. We imposed this constraint because it represents the only case in which the cross-sectional assessment that $c' = 0$ is accurate. Recall that when these stabilities are not equal to one another, the estimation of c' is substantially biased (see Tables 3 and 4). Although this inaccuracy might happen to counterbalance the discrepancies shown in Figure 6, it is also quite possible that this additional discrepancy will serve to increase the bias even further. (Of course, the specific patterns shown in Figure 6

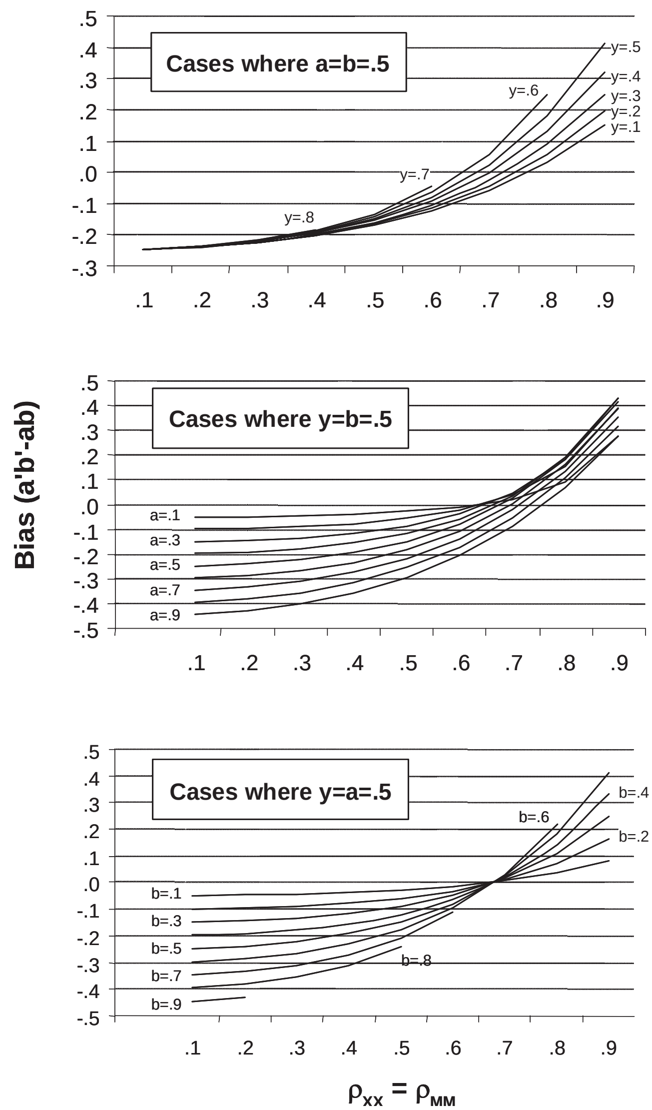

Figure 6. Bias in cross-sectional indirect effect index ($a'b'$). a = direct effect of $X \rightarrow M$; b = direct effect of $M \rightarrow Y$; y = autoregressive effect of Y; a' = cross-sectional direct effect of $X \rightarrow M$; b' = cross-sectional direct effect of $M \rightarrow Y$; ρ_{XX} = stability of X; ρ_{MM} = stability of M.

might not hold if c' were nonzero.) In general, Equation 13 and Figure 6 together show that cross-sectional values of indirect effects are often substantially different from corresponding longitudinal indirect effects, even in very large samples.

Cross-Sectional Estimation of Proportion of Total Effect Mediated by *M*: Autoregressive Model

Suppose some variable M has been found to mediate the relation between two other variables X and Y. An interesting follow-up question is typically to ascertain what proportion of the total effect of X on Y is mediated by the variable M (Shrout & Bolger, 2002). In a cross-sectional analysis such as depicted in Figure 1, the proportion of the total effect that is mediated by M is written as P_M and can be expressed as

$$P_M = \frac{a'b'}{\rho_{XY}} \,. \tag{14}$$

Shrout and Bolger (2002) pointed out that an important advantage of the index P_M is that it provides a continuous measure of the strength of mediation instead of reducing mediation to a dichotomous yes–no decision. In this sense, P_M is an effect size index and conveys information about the magnitude of the mediated effect. MacKinnon, Warsi, and Dwyer (1995) found that large sample sizes (e.g., 500 or more participants) are usually required to obtain stable estimates of the population value of P_M. Shrout and Bolger described a bootstrap procedure for obtaining standard errors and confidence intervals for P_M. Together, these two articles demonstrate the hazards of overinterpreting values of P_M unless point estimates are accompanied by corresponding confidence intervals, especially in small samples. However, an even more basic question is the extent to which a cross-sectional estimate of P_M is likely to be biased, a problem not solved by forming a confidence interval, even in large samples.

How accurate is the cross-sectional index P_M shown in Equation 14 as an index of longitudinal mediation? To answer this question, notice that M completely mediates the relation between X and Y according to the longitudinal model of Figure 5. Thus, from the longitudinal perspective, the true value of P_M equals 1.0. It then follows that the magnitude of the cross-sectional proportional index to the corresponding longitudinal proportional index is simply given by the ratio on the right side of Equation 14:

$$\frac{P_M(\textit{cross-sectional})}{P_M(\textit{longitudinal})} = \frac{a'b'}{\rho_{XY}}, \tag{15}$$

where ρ_{XY} is the correlation between X and Y at a fixed point in time and thus can be denoted as $\rho_{X_tY_t}$.

Equation 15 can be rewritten by substituting for $a'b'$ from Equation B1 in Appendix B, yielding

$$\frac{P_M(\textit{cross-sectional})}{P_M(\textit{longitudinal})} = 1 - \frac{c'}{\rho_{X_tY_t}}. \tag{16}$$

Substituting for c' from Equation 9 and for $\rho_{X_tY_t}$ from Appendix A and subsequently rearranging terms yields

$$\frac{P_M(\textit{cross-sectional})}{P_M(\textit{longitudinal})} = 1 - \frac{\rho_{X_tX_{t-1}} - \rho_{M_tM_{t-1}}}{x(1-my)(1-\rho^2_{X_tM_t})}, \tag{17}$$

as long as a, b, and x are all nonzero. Equation 17 shows that the cross-sectional index P_M will always differ from the longitudinal index P_M except in the special case where X and M are equally stable. Furthermore, when X is more stable than M, the cross-sectional index will be smaller than the longitudinal index. Conversely, when X is less stable than M, the cross-sectional index will be larger than the longitudinal index.

Figure 7 provides an indication of the possible magnitude of the discrepancy between the cross-sectional and longitudinal indices. The x-axis of the figure depicts the parameter x, which is the stability of X. The y-axis depicts Equation 17, expressed as a percentage. Thus, the horizontal line at a height of 100 represents situations where the cross-sectional and longitudinal indices are equivalent to one another. Points above this line reflect situations where the cross-sectional index is larger than the longitudinal index, whereas points below this line reflect situations where the cross-sectional index is smaller than the longitudinal index. The figure shows the percentage for two arbitrary but reasonable sets of parameter values. In particular, when $a = 0.30$, $m = 0.69$, and $y = 0.60$, the cross-sectional index tends to be larger than the longitudinal index. For example, when the stability of X is around .50, the cross-sectional index is nearly twice as large as the longitudinal index. The cross-sectional index remains larger than the longitudinal index until the stability of X exceeds .91, at which point the percentage begins to drop rapidly. When the stability of X reaches its maximum value of 1.00, the cross-sectional index is less than half as large as the longitudinal index for this configuration of parameter values.

A second set of parameter values shows a very different curve (see Figure 7). Specifically, when $a = 0.50$, $m = 0.33$, and $y = 0.50$, the cross-sectional index is always less than the longitudinal index for stability values of X between .50 and 1.00. When the stability of X is near .50, the cross-sectional index is only slightly less than the longitudinal index; however, the ratio steadily declines for more stable values of X, reaching a point where the cross-sectional index is less than 20% of the longitudinal index as X approaches perfect stability. In general, Equation 17, together with Figure 7, shows that even in very large samples, a cross-sectional index of the proportion of total effect mediated by a variable M can be either much larger or much smaller than the corresponding longitudinal index. Furthermore, the

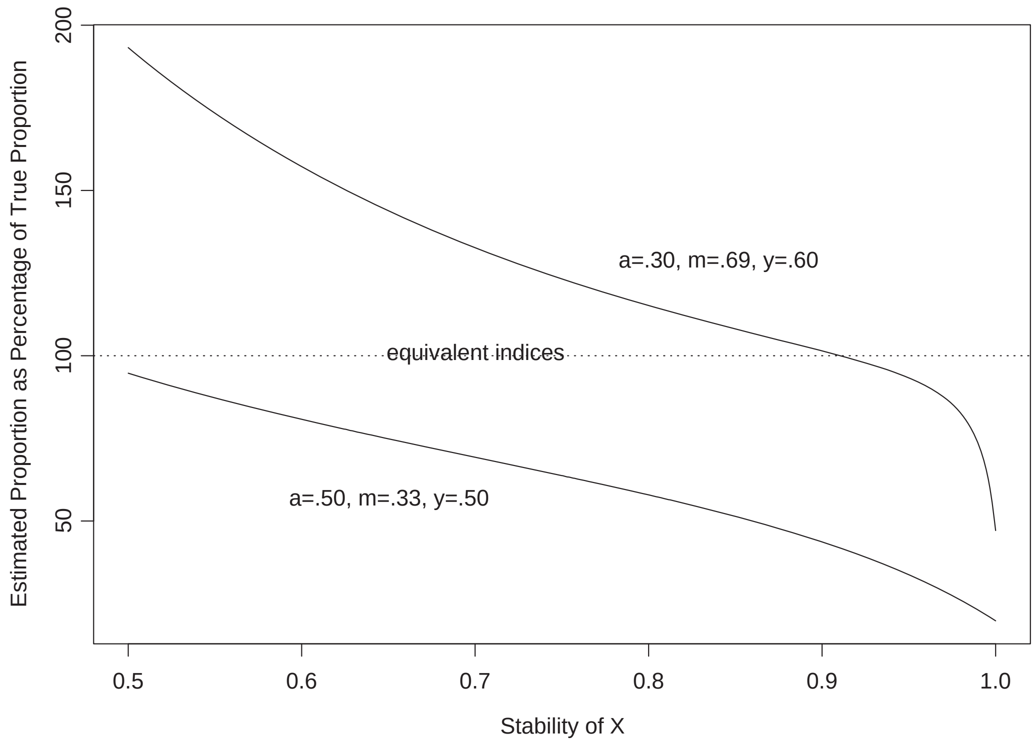

Figure 7. Estimated P_M (proportion of the total effect of X on Y that is mediated by M) as a percentage of actual P_M for selected parameter values. X = independent variable; M = mediator; Y = dependent variable; a = direct effect of $X \rightarrow M$; m = autoregressive effect of M; y = autoregressive effect of Y.

cross-sectional data do not provide enough information to evaluate whether the cross-sectional index is likely to be too small or too large, so the direction of bias is unknown in the absence of longitudinal data.

A Random Effects Model of Mediation

We have seen that cross-sectional analyses can produce very misleading conclusions regarding mediation when the true nature of change follows an autoregressive model. However, the autoregressive model is only one plausible model of change. We now turn our attention to a rival model, namely, the random effects model. There are many different specific models we might consider because random effects models are actually an entire class of models. For this reason, we rely on the specific random effects model of mediation proposed by Kenny et al. (2003) in which *X, M,* and *Y* change over time. In doing so, we also want to acknowledge that Cheong, MacKinnon, and Khoo (2001, 2003) have developed an appropriate random effects model for a situation where *X* represents a fixed intervention. As in the case of the autoregressive model, we consider the special case where the mediator *M* completely mediates the effect of *X* on *Y*; however, we note that the model developed by Kenny et al. (2003) also allows for partial mediation.

In the case of complete mediation, Kenny et al.'s (2003) model stipulates that *X*, *M*, and *Y* for individual *i* at time *t* are related as follows:

$$M_{it} = d_{1i} + a_i X_{it} + e_{it}, \text{ and} \tag{18}$$

$$Y_{it} = d_{2i} + b_i M_{it} + f_{it}. \tag{19}$$

To understand this model more completely, let us borrow an example originally presented by Kenny et al. Suppose that *X* is an indicator variable reflecting the occurrence or absence of a specific stressor for an individual on a given day. (For our purposes, it does not matter whether *X* is categorical or continuous, so *X* could also reflect level of stress on a given day.) Further suppose that *M* reflects an individual's level of coping on that day and *Y* reflects the individual's mood that day. In a longitudinal design, all three variables would be measured over a period of days for a group of individuals.

According to Equations 18 and 19, an individual's coping (*M*) on any given day will depend in part on whether the individual experienced the specific stressor (*X*) on that day.

Similarly, that individual's mood will depend in part on his or her level of coping on that day. The fact that X (stress) does not appear as an explanatory variable for Y (mood) in Equation 19 reflects the fact that M completely mediates the relation between X and Y in this version of the model. Of course, rarely will stress perfectly explain coping, and rarely will coping perfectly explain mood. Unexplained effects are represented by the error terms e_{it} and f_{it}.

Several characteristics of this model are especially important for our purposes. First, notice that scores on X are allowed to change over time. Equation 18 shows that such changes on X will translate into corresponding changes on M to the extent that the a_i parameter is nonzero. Similarly, from Equation 19, changes on M will result in changes on Y to the extent that the b_i parameter is nonzero. Thus, even though time is not an explicit term in either equation, the model is in fact a model of change insofar as X is allowed to change over time (as reflected by the t subscript associated with X), M is dependent on X, and Y is dependent on M.[6] Second, notice that the extent to which M depends on X and the extent to which Y depends on M can vary across individuals, as reflected by the presence of an i subscript for the a_i and b_i parameters. Third, notice the presence of two additional random effects, namely, d_{1i} and d_{2i}. These random effects allow M and Y to exhibit stability over time.

We show below that d_{1i} and d_{2i} play an especially important role in our derivations, so we want to take a moment now to consider their meaning. To understand the role played by these terms, we first consider d_{1i} in Equation 18. Notice that the presence of an i subscript for this parameter indicates that it is a random effect allowed to vary over individuals. However, the absence of a t subscript implies that this effect does not vary over time and thus is a perfectly stable individual-differences variable. Suppose for a moment that every individual had the same value for the d_{1i} parameter (i.e., suppose that the variance of this random effect parameter were zero). This would imply that every person experiencing the stressor should have the same level of coping except for random error. Similarly, if every person had the same value of d_{2i}, all individuals with a specific level of coping should have the same level of mood except for random error. In reality, however, consistent individual differences in coping almost certainly exist even holding the stressor constant. Likewise, consistent individual differences in mood almost certainly exist even for persons who exhibit the same level of coping. Allowing d_{1i} and d_{2i} to vary over individuals allows for such consistent individual differences.

It may be useful to pause for a moment to consider how the random effects model allows for stability over time. As we have just seen, the d_{1i} and d_{2i} random effects reflect consistent individual differences. In essence, these random effects represent baseline levels of M and Y. For example, as the presence of the stressor X fluctuates from one day to the next, a given person's level of coping (M) similarly fluctuates as a function of the a_i parameter. The critical point is that this fluctuation occurs around the d_{1i} value, and the level of d_{1i} generally varies from person to person. In other words, some individuals simply cope better than others either in the presence or absence of the specific stressor X. To the extent that this is true, daily M scores are fluctuating around each person's baseline level of coping. As a consequence, M will exhibit some degree of stability in the sense that scores across individuals at one time point will tend to correlate with scores at another time point because each individual has his or her constant baseline level around which scores fluctuate over time. A similar argument applies for Y.

Estimating the Cross-Sectional Direct Effect of X on Y: Random Effects Model

Suppose that the random effects model of the previous section depicts the true nature of the process whereby M completely mediates the effect of X on Y. To what extent is it possible to capture this process accurately with cross-sectional data? We pursue the answer to this question by assuming that the mediational process follows the Kenny et al. (2003) model of mediation and then determining how well a cross-sectional approach reveals the true longitudinal mediation process.

Suppose a researcher collects cross-sectional data on X, M, and Y at some fixed point in time. Appendix C shows that when complete mediation occurs in the Kenny et al. (2003) model, the correlations among X, M, and Y at any point in time t can be written in terms of the underlying model parameters as follows:

$$\rho_{XM} = a + \sigma_{Xd_1}, \tag{20}$$

$$\rho_{MY} = b + \sigma_{d_1 d_2} + a\sigma_{Xd_2}, \text{ and} \tag{21}$$

$$\rho_{XY} = ab + \sigma_{Xd_2} + b\sigma_{Xd_1}. \tag{22}$$

The direct effect of X on Y in a cross-sectional mediational model equals the standardized regression coefficient for X in a regression model where Y is predicted from both X and M, which can be written as

[6] This model specifies contemporaneous effects of X on M and of M on Y. In some situations, such as daily stress (X), daily coping (M), and daily mood (Y), as discussed by Kenny et al. (2003), this model seems reasonable. In other situations, lagged effects might be deemed more appropriate. An extension of our derivations could consider a cross-sectional analysis of such a lagged longitudinal model. Even more stringent conditions would need to hold before a cross-sectional analysis would yield unbiased estimates of parameters in a lagged longitudinal model.

$$\beta_{YX \cdot M} = \frac{\rho_{XY} - \rho_{XM}\rho_{MY}}{1 - \rho_{XM}^2}. \tag{23}$$

The population value of this direct effect can be found by substituting expressions for the correlations from Equations 20 through 22 into Equation 23, which, after straightforward algebra, reduces to

$$\beta_{YX \cdot M} = \frac{\sigma_{Xd_2} - (\sigma_{d_1d_2} + a\sigma_{Xd_2})(a + \sigma_{Xd_1})}{1 - \rho_{XM}^2}. \tag{24}$$

However, keep in mind that Equations 20 through 22 are based on a model of complete mediation as reflected by the model formulation shown in Equations 18 and 19. Thus, the true value of the direct effect is zero. Consequently, the expression in Equation 24 represents the bias in the cross-sectional estimate of the true direct effect in the random effects model.

To understand the bias shown in Equation 24, we need to probe the meaning of three covariance terms, $\sigma_{d_1d_2}$, σ_{Xd_1}, and σ_{Xd_2}. Recall that the random effect d_{1i} represents the baseline level of M (controlling for X) for individual i, whereas the random effect d_{2i} represents the baseline level of Y (controlling for M) for individual i. Thus, Equation 24 shows that the magnitude of bias in the direct effect depends on the covariances among these random effects. Would these random effects typically be correlated with one another? To answer this question, let us return to the example where X represents the presence of a stressor, M is coping, and Y is mood. Then, the random effect d_{1i} represents the baseline level of daily coping (controlling for the stressor) for individual i. Similarly, the random effect d_{2i} represents the adjusted baseline level of daily mood (controlling for coping) for individual i. In other words, each person has some adjusted baseline level of daily mood, taking into account the influence that daily coping has on daily mood for each person. The covariance $\sigma_{d_1d_2}$ will be nonzero if individuals with higher adjusted baseline levels of coping have different adjusted baseline levels of mood than individuals with lower adjusted baseline levels of coping.

The covariance of X with d_1 reflects the extent to which individuals with higher levels of the stressor (X) tend to have different adjusted baseline levels of coping (M) than individuals with lower levels of the stressor. Similarly, the covariance of X with d_2 reflects the extent to which individuals with higher levels of the stressor tend to have different adjusted baseline levels of mood than individuals with lower levels of the stressor.

Only under extremely limited conditions will a cross-sectional analysis provide an unbiased estimate of the true direct effect from the random effects model. The bias shown in Equation 24 will equal zero if and only if

$$\sigma_{Xd_2} - (\sigma_{d_1d_2} + a\sigma_{Xd_2})(a + \sigma_{Xd_1}) = 0. \tag{25}$$

In practice, Equation 25 will hold under remarkably few conditions. The most straightforward case where the bias will equal zero is if both covariances $\sigma_{d_1d_2}$ and σ_{Xd_2} happen to equal zero, but it is difficult to develop a convincing rationale for such an expectation. In particular, these covariances will equal zero only if (a) all influences on M other than X are uncorrelated with all influences on Y other than M and (b) all influences on Y other than M are uncorrelated with X.[7]

A potentially important simplification occurs in the special case where X is randomly assigned to each individual, as Kenny et al. (2003) suggested is ideal. To understand how random assignment might work here, it is important to realize that by design, X varies within a person. Thus, random assignment of X here would mean that on certain randomly selected days, X would take on one value, whereas on other randomly selected days, X would take on a different value. For example, a high stress induction might be introduced on certain random days but not on others. In any event, when X is randomly assigned, X cannot covary in the population with either d_1 or d_2. Thus, for there to be no bias in the cross-sectional design, d_1 and d_2 must be uncorrelated. Although this is a weaker assumption than the one underlying the more general case, it will nevertheless be a difficult one for most researchers to justify, thus still limiting the value of the cross-sectional analysis.

Equation 25 suggests that cross-sectional estimates of the direct effect will typically be biased. Equation 24 provides the mathematical expression for the magnitude of bias. Nevertheless, it is difficult to judge from these equations how large the bias is likely to be in practical terms. Figure 8 shows the magnitude of bias in the cross-sectional direct effect for certain fixed values of model parameters. Specifically, this figure is based on parameter values of $a = 0.60$ and $b = 0.50$. Furthermore, the standard deviations of both d_1 and d_2 are assumed to equal 0.50. For simplicity, ρ_{Xd_1} is assumed to be equal to ρ_{Xd_2}. The x-axis of the figure represents this common correlation between X and each of the d variables. Finally, two curves appear in the figure, one corresponding to a situation where d_1 and d_2 are uncorrelated and another corresponding to a situation where they correlate 0.50 with one another.

Several practical implications emerge from Figure 8. First, the magnitude of bias in the cross-sectional direct effect parameter can be substantial, sometimes reaching values that would typically be interpreted as a large effect size even when the true longitudinal parameter equals zero.

[7] A more general formulation of the random effects model could include time as an explicit predictor of both M and Y. Appendix D derives an expression comparable to Equation 25, demonstrating the conditions under which the cross-sectional estimate of the direct effect will be unbiased in this more general formulation.

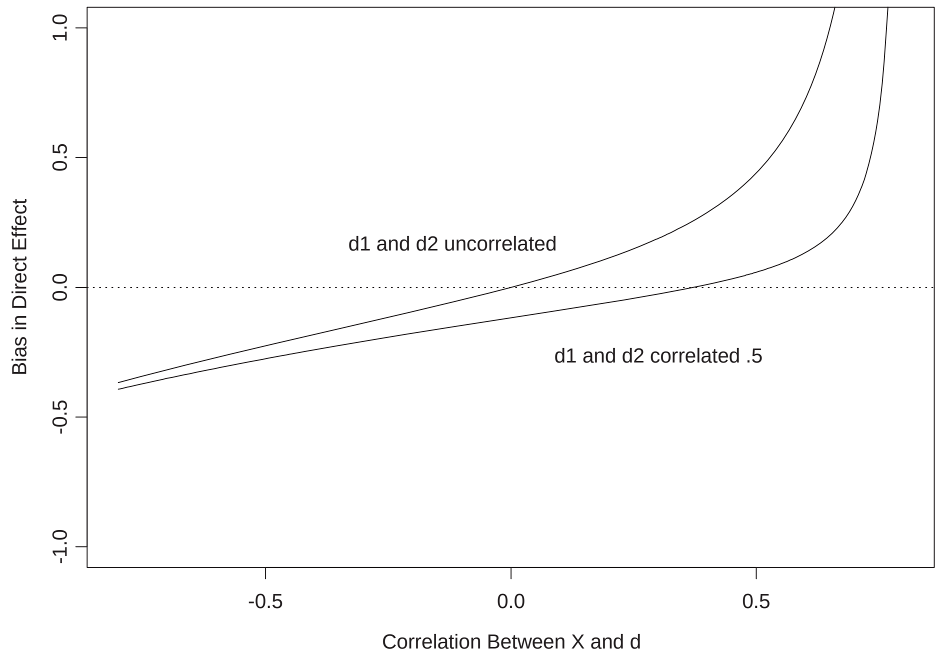

Figure 8. Bias in cross-sectional direct effect when $a = 0.60$, $b = 0.50$, $\sigma_{d_1} = \sigma_{d_2} = 0.50$, and $\rho_{Xd_1} = \rho_{Xd_2}$. X = independent variable; d = random effect; a = direct effect of $X \rightarrow M$; b = direct effect of $M \rightarrow Y$.

Second, the direction of bias can be either positive or negative. In other words, the cross-sectional estimate can be either too large or too small depending primarily on the magnitude of the correlations of X with d_1 and d_2. Unfortunately, this makes meaningful interpretation especially complicated because, with cross-sectional data, there is no way to know whether direct effect estimates will generally be too small or too large. Third, the correlation between d_1 and d_2 has relatively little effect on the bias until X becomes positively correlated with d_1 and d_2, at which point the correlation between the d variables becomes an important determinant of the bias. Fourth, even though Figure 8 pertains to a specific set of parameter values, the basic point is more general: Substantial bias can exist in cross-sectional estimates of direct effects when the true mediational process follows the longitudinal random effects model (represented in Equations 18 and 19).

Estimating the Cross-Sectional Indirect Effect of X on Y Through M: Random Effects Model

The previous section has shown that stringent assumptions must be met for a cross-sectional design to yield unbiased estimates of longitudinal mediation based on the random effects model. We now turn our attention to the corresponding indirect effect. A distinguishing characteristic of the random effects model is that each individual is allowed to have a unique indirect effect because the effect of X on M is allowed to vary over individuals, as is the effect of M on Y. As a consequence, Kenny et al. (2003) showed that the average indirect effect is not necessarily equal to ab because

$$E(a_i b_i) = ab + \sigma_{ab}. \tag{26}$$

Thus, assessing bias in a cross-sectional design requires comparing the expression in Equation 26 with the expected value of the indirect effect as assessed in a cross-sectional design. Appendix E shows that this bias can be written as

$$bias = (a + \sigma_{Xd_1}) \times \left[\frac{(b + \sigma_{d_1 d_2} + a\sigma_{Xd_2}) - (a + \sigma_{Xd_1})(ab + \sigma_{Xd_2} + b\sigma_{Xd_1})}{1 - \rho^2_{XM}} \right] - (ab + \sigma_{ab}). \tag{27}$$

Recall that the cross-sectional estimate of the direct effect is unbiased if the covariance terms $\sigma_{d_1 d_2}$ and σ_{Xd_2} equal zero. Even if both of these terms and the additional covariance term σ_{Xd_1} all equal zero, the bias in the indirect effect simplifies not to zero but instead to

$$bias = -\sigma_{ab}. \tag{28}$$

Thus, even under these conditions, the bias will still be nonzero to the extent that there is some correlation between the a_i and b_i random effects.

Figure 9 shows the magnitude of bias in the cross-sectional indirect effect for certain fixed values of model parameters. Specifically, all parameters are assumed to have the same values as in Figure 8. As shown in Equations 26 through 28, another parameter that must be considered when evaluating bias in indirect effects is the covariance between a_i and b_i. For simplicity, this covariance is assumed to equal zero in Figure 8. However, inspection of Equation 27 shows that the influence of a nonzero covariance would simply shift the bias upward or downward by an amount exactly equal to the covariance itself. Of course, if the covariance between a_i and b_i were substantial, this could have a sizable effect on the bias of the cross-sectional indirect effect. In some situations, such a covariance could counteract the bias depicted in Figure 9, whereas in other situations, it could exacerbate the bias depicted in the figure.

Several practical implications emerge from Figure 9. First, the magnitude of bias in the cross-sectional indirect effect parameter can be substantial, often reaching values that would typically be interpreted as corresponding to a medium effect size and occasionally exceeding values that would usually be considered to be large. Second, as we have seen for the direct effect, the direction of bias in the indirect effect can be either positive or negative. In other words, the cross-sectional estimate of the indirect effect can be either too large or too small depending primarily on the magnitude of the correlation of X with d_1 and d_2. Unfortunately, this again makes meaningful interpretation of cross-sectional results especially complicated because there is no general way to know whether indirect effect estimates obtained from cross-sectional analyses will be too small or too large. Third, the correlation between d_1 and d_2 has relatively little effect on the bias in the indirect effect until X becomes positively correlated with d_1 and d_2, at which point the degree of correlation between the d variables becomes an important determinant of the bias. Fourth, comparing the pattern of bias in Figure 9 with the pattern shown earlier in Figure 8 reveals that (for certain configurations of parameter values) both the indirect effect and the direct effect may be substantially overestimated. At first glance, this may seem counterintuitive because the total effect is partitioned into these two components. However, implicit in these two figures is the point that the total effect in the cross-sectional design may bear little resemblance to the longitudinal total effect. Thus, the total amount being partitioned can be much larger in the cross-sectional design, leading to an overestimation of both the direct effect and the indirect effect. Fifth, it should be kept in mind that this figure pertains to a specific set of parameter values. Nevertheless, the basic point is more general. Namely, substantial bias can and typically will contaminate cross-sectional estimates of indirect effects when the true mediational process follows the longitudinal random effects model represented in Equations 18 and 19.

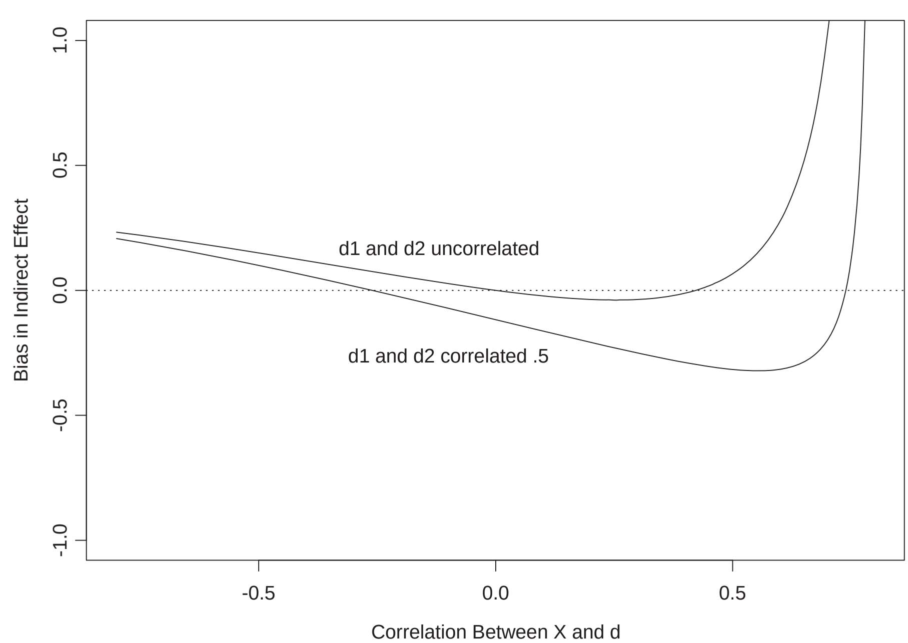

Figure 9. Bias in cross-sectional indirect effect when $a = 0.60$, $b = 0.50$, $\sigma_{d_1} = \sigma_{d_2} = 0.50$, $\rho_{Xd_1} = \rho_{Xd_2}$, and $\sigma_{ab} = 0$. X = independent variable; d = random effect; a = direct effect of $X \rightarrow M$; b = direct effect of $M \rightarrow Y$.

Cross-Sectional Estimation of Proportion of Total Effect Mediated by *M*: Random Effects Model

As discussed in the context of the autoregressive model, an interesting follow-up question in mediation analysis involves ascertaining what proportion of the total effect of X on Y is mediated by the variable M. Of particular interest is the extent to which a cross-sectional analysis can be relied upon to reflect the true proportion if the longitudinal process conforms to the random effects model shown in Equations 18 and 19.

Under this random effects model, the ratio of the cross-sectional proportional index to the corresponding longitudinal proportional index is

$$\frac{P_M(\textit{cross-sectional})}{P_M(\textit{longitudinal})} = \frac{\beta_{MX}\beta_{YM\cdot X}}{\rho_{XY}}. \tag{29}$$

Substituting from Equations C6, E3, and C11 in the appendixes into Equation 29 yields

$$\frac{P_M(\textit{cross-sectional})}{P_M(\textit{longitudinal})} = \frac{(a + \sigma_{Xd_t})}{(ab + \sigma_{Xd_2} + b\sigma_{Xd_t})} \times \frac{(b + \sigma_{d_1d_2} + a\sigma_{Xd_2}) - (a + \sigma_{Xd_1})(ab + \sigma_{Xd_2} + b\sigma_{Xd_1})}{1 - \rho^2_{xM}}. \tag{30}$$

In general, the cross-sectional proportion index will differ from the longitudinal index. The only plausible exception is the case where all three previously mentioned covariances, namely, $\sigma_{d_1d_2}$, σ_{Xd_1}, and σ_{Xd_2}, equal zero. In this special case, the expression shown in Equation 30 simplifies to a value of 1.00, implying that the cross-sectional proportion index will be identical to the longitudinal index in the population. However, when one or more of these covariances are nonzero, the cross-sectional index is generally not the same as the longitudinal index. Figure 10 shows the discrepancy between the cross-sectional index and the longitudinal index for a fixed but reasonable set of values of these covariance parameters. Specifically, all parameters are once again assumed to have the same values as in Figure 8.

Several practical implications emerge from Figure 10. First, the magnitude of bias in the cross-sectional proportional index can be substantial. Second, as we have shown to be true of the direct effect and the indirect effect, the direction of bias can be either positive or negative. In other words, the cross-sectional estimate of the proportion of the total effect that is mediated can be either too large or too small. Once again, this makes meaningful interpretation of cross-sectional results especially complicated. Third, the bias in the cross-sectional index is sensitive to the correlation of X with d_1 and d_2 as well as to the correlation between d_1 and d_2. Larger values of the correlation of X with d_1 and d_2 lead to smaller relative sizes of the cross-sectional index.

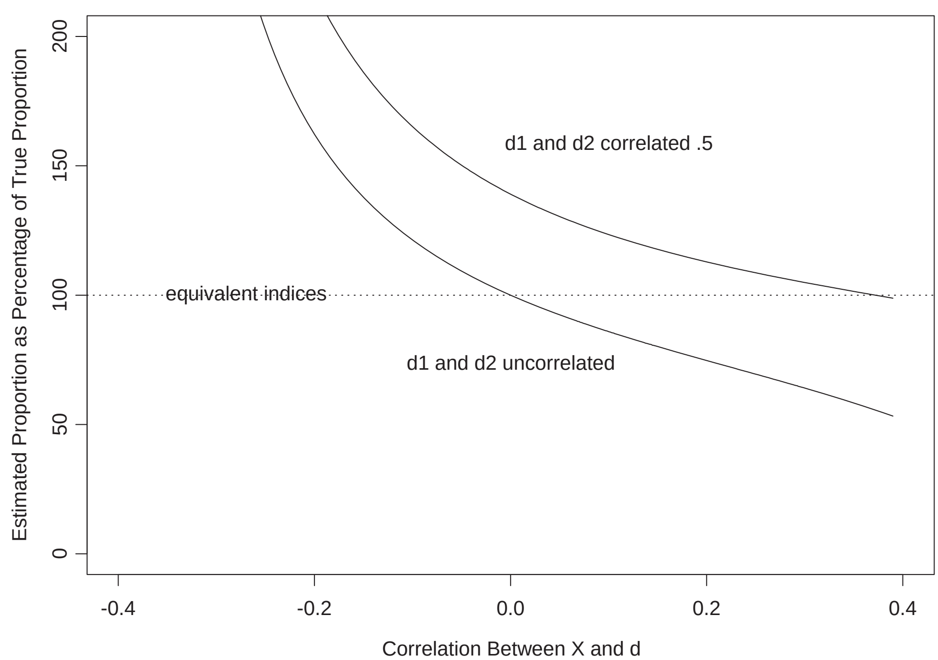

Figure 10. Estimated P_M (proportion of the total effect of X on Y that is mediated by M) as a percentage of actual P_M when $a = 0.60$, $b = 0.50$, $\sigma_{d_1} = \sigma_{d_2} = 0.50$, and $\rho_{Xd_1} = \rho_{Xd_2}$. X = independent variable; M = mediator; Y = dependent variable; d = random effect; a = direct effect of $X \rightarrow M$; b = direct effect of $M \rightarrow Y$.

However, whether this reduces or increases the bias in the cross-sectional index depends in no small part on the correlation between d_1 and d_2. Fourth, although the figure represents a specific set of parameter values, the more general point remains: Substantial bias can exist in cross-sectional estimates of the proportion of the total effect that is mediated when the true mediational process follows the longitudinal random effects model represented in Equations 18 and 19.

Conclusion

The major conclusion of this article is that cross-sectional examination of mediation will typically generate biased estimates of longitudinal mediation parameters even under the ideal situation when mediation is complete. Furthermore, the magnitude of bias can be substantial. This conclusion holds for two rather different types of longitudinal models, namely, an autoregressive model and a random effects model.

Although bias in cross-sectional analyses of mediation is likely regardless of which longitudinal model accurately reflects the true longitudinal nature of mediation, the exact pattern of bias depends on the specific longitudinal model. For the autoregressive model, three sets of specific conclusions emerge. First, estimates of the cross-sectional direct effect of predictor X on outcome Y are positively biased when X is more stable than the mediator M; conversely, estimates of the cross-sectional direct effect of X on Y are negatively biased when M is more stable than X. The magnitude of this bias can be substantial. Second, estimates of the cross-sectional indirect effect of X on Y through M are also substantially biased under a wide range of conditions. When X and M are relatively unstable, the cross-sectional indirect effect can be substantially negatively biased; however, when X and M are relatively stable, the cross-sectional indirect effect can be substantially positively biased. Third, cross-sectional estimates of the proportion of the total effect mediated by M also evince substantial positive or negative bias relative to the corresponding longitudinal index. The direction of this bias depends upon the relative stabilities of X and M. Given previous work on the relation between cross-sectional and longitudinal models (e.g., Gollob & Reichardt, 1985, 1987, 1991), it should perhaps come as no surprise that cross-sectional designs and analyses cannot generally be counted on as faithful representations of longitudinal processes. Nevertheless, the degree of bias revealed in the current article suggests that cross-sectional tests of longitudinal mediation processes have extremely limited applicability.

The fundamental reason for the inability of the cross-sectional model to capture longitudinal processes as reflected by the autoregressive model is its failure to allow for autoregressive effects of M and Y across time. That is, cross-sectional models do not allow statistical control for prior M or prior Y. As such, the cross-sectional model is misspecified, and parameter estimates are generally biased. As Reichardt and Gollob (1986) pointed out, the cross-sectional model is also misspecified because it fails to allow for causation to occur over time and instead presumes that X at time t causes M at the same time t. The preferred solution to this problem is to include prior measures of X, M, and Y in the model to allow for autoregressive effects and time lags in presumed causal effects. An alternative approach is Gollob and Reichardt's (1987) latent longitudinal analysis when only cross-sectional data are available. Unfortunately, this approach requires tenuous assumptions or substantial prior knowledge about longitudinal parameters that underlie the cross-sectional data.

Cross-sectional analyses will also typically fail to represent longitudinal processes as reflected by a longitudinal random effects model. To the extent that baseline levels of the mediator and the outcome vary across individuals and correlate either with one another or with levels of the presumed cause (X), bias will emerge in cross-sectional estimates of the direct effect, the indirect effect, and the proportion of the total effect that is mediated by the mediator. The magnitude of bias can be substantial, and to make matters worse, the direction of bias is unknown unless strong assumptions can be made about various covariances among the random effects parameters. As in the autoregressive model, the failure of cross-sectional analysis stems from its inability to model stable relations between variables over time. In other words, the cross-sectional analysis is unable to determine the extent to which correlations between measures reflect an influence of one measure on another over time or instead reflect ongoing stable relations between measures.

Conspicuous by its absence from this article is any discussion of hypothesis testing. We have not considered hypothesis testing here because the bias inherent in cross-sectional designs has disastrous consequences for hypothesis tests. In particular, the fact that cross-sectional parameters generally differ from corresponding longitudinal parameters implies that hypothesis tests based on cross-sectional data will also be biased (cf. Casella & Berger, 2002). As a consequence, when the population value of a longitudinal parameter is zero, we have seen that the corresponding cross-sectional parameter can be very different from zero. In this case, it is entirely possible for the rejection rate of the hypothesis test to approach 1.00 even though the longitudinal parameter exactly equals zero (i.e., the longitudinal null hypothesis is true). Type I errors become essentially inevitable, especially in large samples. Similarly, tests of non-null hypotheses can have very low power because a cross-sectional parameter value close to zero can correspond to a sizable longitudinal parameter value. In such a case, statistical power can be very low even though the true longitudinal parameter is very different from zero. Unfortunately, avoiding hypothesis tests by relying on confidence intervals fares no better. It is possible that the true coverage rate of a 95% cross-sectional confidence interval for a longitudinal parameter approaches a value of 0%. In fact, larger sample sizes will typically lead to lower coverage probabilities because the interval hones in on a biased estimate of the true longitudinal parameter. What appears to be a very accurate interval (because of a large sample size) may in fact have an upper bound and a lower bound neither of which is remotely close to the true longitudinal parameter value. The important practical point here is that the substantial bias that typically exists in cross-sectional analyses of mediation can render *p* values or confidence intervals obtained from cross-sectional data essentially meaningless.

Longitudinal designs offer additional advantages beyond the ability to eliminate bias in parameter estimates. Perhaps most importantly, as MacKinnon et al. (2002) emphasized, longitudinal designs can yield information about temporal precedence and thus allow examination of which variables are causes and which variables are effects. For example, it may be the case that although maternal depression contributes to child depression, the opposite is also true. Longitudinal designs are especially well suited to examine such complex causal relations.

Several limitations of the current work suggest avenues for future research. First, we have focused only on what many researchers may regard as the ideal (or hoped-for) situation: one in which there is complete mediation. Our results show that cross-sectional analyses of processes involving complete longitudinal mediation cannot generally be trusted to yield accurate estimates of true underlying longitudinal processes. A more complex situation arises when longitudinal mediation is incomplete (i.e., when *M* is controlled, the longitudinal direct effect of *X* on *Y* does not go to zero). The magnitude of bias in cross-sectional analyses remains to be explored in the more complicated case of partial longitudinal mediation.

A second limitation is that the current article does not examine the role of time-lag duration in the design and analysis of longitudinal studies. As has been discussed elsewhere (e.g., Cole & Maxwell, 2003; Gollob & Reichardt, 1985, 1987, 1991), estimates of effects in longitudinal models can change greatly depending on the chosen time lag. In this respect, continuous time models (e.g., Boker, Neale, & Rausch, 2004; Oud & Jansen, 2000) offer an interesting alternative because parameter values are unaffected by choice of time lag between adjacent measurement occasions.

A third limitation is that we constrained the current article to focus on mediation when *X*, *M*, and *Y* are all changing over time. Important special cases exist in which *X* might be fixed in time. In experimental designs or therapy outcome studies, where participants are assigned to treatment or control conditions, investigators are often interested in the potential mediators of the manipulation or treatment. Close examination of mediation in such designs is clearly warranted.

In summary, we find that cross-sectional approaches to longitudinal mediation can substantially over- or underestimate longitudinal effects even under the ideal conditions where mediation is complete, longitudinal parameter estimates are completely stable, and sample size is very large. We anticipate that similar problems will emerge under less ideal circumstances, such as partial mediation. We urge researchers interested in mediational processes to include multiple waves of data in their designs and analyses, and we call upon methodologists to continue development of appropriate models for understanding how psychological processes unfold over time.

References

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.

Bohrnstedt, G. W., & Goldberger, A. S. (1969). On the exact covariance of products of random variables. *Journal of the American Statistical Association, 64,* 1439–1442.

Boker, S. M., Neale, M. C., & Rausch, J. R. (2004). Latent differential equation modeling with multivariate multi-occasion indicators. In K. van Montfort, H. Oud, & A. Satorra (Eds.), *Recent developments on structural equations models: Theory and applications* (pp. 151–174). Dordrecht, the Netherlands: Kluwer Academic.

Casella, G., & Berger, R. L. (2002). *Statistical inference* (2nd ed.). Pacific Grove, CA: Duxbury.

Cheong, J., MacKinnon, D. P., & Khoo, S. T. (2001). A latent growth modeling approach to mediation analysis. In L. M. Collins & A. G. Sayer (Eds.), *New methods for the analysis of change* (pp. 390–392). Washington, DC: American Psychological Association.

Cheong, J., MacKinnon, D. P., & Khoo, S. T. (2003). Investigation of mediational processes using parallel process latent growth curve modeling. *Structural Equation Modeling, 10,* 238–262.

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed.). Mahwah, NJ: Erlbaum.

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.

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

Gollob, H. F., & Reichardt, C. S. (1985). Building time lags into causal models of cross-sectional data. *Proceedings of the Social Statistics Section of the American Statistical Association, 28,* 165–170.

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: Recent advances, unanswered questions, future directions* (pp. 243–259). Washington, DC: American Psychological Association.

Kenny, D. A. (1979). *Correlation and causality.* New York: Wiley.

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

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

Lord, F. M. (1967). A paradox in the interpretation of group comparisons. *Psychological Bulletin, 68,* 304–305.

Lord, F. M. (1969). Statistical adjustments when comparing preexisting groups. *Psychological Bulletin, 72,* 336–337.

MacCallum, R. C., & Austin, J. T. (2000). Applications of structural equation modeling in psychological research. *Annual Review of Psychology, 51,* 201–226.

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

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

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

Oud, J. H. L., & Jansen, R. A. R. G. (2000). Continuous time state space modeling of panel data by means of SEM. *Psychometrika, 65,* 199–215.

Reichardt, C. S., & Gollob, H. F. (1986). Satisfying the constraints of causal modeling. In W. M. K. Trochim (Ed.), *Advances in quasi-experimental design and analysis* (pp. 91–107). San Francisco: Jossey-Bass.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference.* Boston: Houghton Mifflin.

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

Tein, J.-Y., Sandler, I. N., MacKinnon, D. P., & Wolchik, S. A. (2004). How did it work? Who did it work for? Mediation in the context of a moderated prevention effort for children of divorce. *Journal of Consulting and Clinical Psychology, 72,* 617–624.

Appendix A

Derivation of Correlations From Autoregressive Model

Expression for the *XM* Correlation

From the model depicted in Figure 5, we see that

$$X_t = xX_{t-1} + \varepsilon_{X_t}, \text{ and} \tag{A1}$$

$$M_t = mM_{t-1} + aX_{t-1} + \varepsilon_{M_t}. \tag{A2}$$

The covariance of X_t and M_t can be expressed as

$$C(X_t, M_t) = mxC(X_{t-1}, M_{t-1}) + axC(X_{t-1}, X_{t-1}). \tag{A3}$$

Assuming that all variables are standardized, covariances are equal to correlations, so Equation A3 can be rewritten as

$$\rho_{X_tM_t} = mx\rho_{X_{t-1}M_{t-1}} + ax. \tag{A4}$$

At equilibrium, correlations are equal across waves, which implies that

$$\rho_{X_tM_t} = \rho_{X_{t-1}M_{t-1}}. \tag{A5}$$

Substituting Equation A5 into Equation A4 yields

$$\rho_{X_tM_t} = mx\rho_{X_tM_t} + ax. \tag{A6}$$

Rearranging terms and solving for the correlation between X and M leads to

(*Appendixes continue*)

$$\rho_{X_tM_t} = \frac{ax}{(1 - mx)}. \quad \text{(A7)}$$

Expression for the *XY* Correlation

From the Figure 5 model, we see that

$$Y_t = bM_{t-1} + yY_{t-1} + \varepsilon_{Y_t}, \text{ and} \quad \text{(A8)}$$

$$X_t = xX_{t-1} + \varepsilon_{X_t}. \quad \text{(A9)}$$

The covariance of X_t and Y_t can be expressed as

$$C(X_t, Y_t) = bxC(X_{t-1}, M_{t-1}) + xyC(X_{t-1}, Y_{t-1}). \quad \text{(A10)}$$

Assuming standardized variables and equal correlations across waves, Equation A10 can be rewritten as

$$\rho_{X_tY_t} = bx\rho_{X_tM_t} + xy\rho_{X_tY_t}. \quad \text{(A11)}$$

Rearranging terms and substituting from Equation A7 yields

$$\rho_{X_tY_t} = \frac{abx^2}{(1 - mx)(1 - xy)}. \quad \text{(A12)}$$

Expression for the *MY* Correlation

From the Figure 5 model, we see that

$$M_t = mM_{t-1} + aX_{t-1} + \varepsilon_{M_t}, \text{ and} \quad \text{(A13)}$$

$$Y_t = bM_{t-1} + yY_{t-1} + \varepsilon_{Y_t}. \quad \text{(A14)}$$

It follows that the covariance of M_t and Y_t can be expressed as

$$C(M_t, Y_t) = bmC(M_{t-1}, M_{t-1}) + abC(X_{t-1}, M_{t-1}) + myC(M_{t-1}, Y_{t-1}) + ayC(X_{t-1}, Y_{t-1}). \quad \text{(A15)}$$

Assuming standardized variables and equal correlations across waves, Equation A15 can be rewritten as

$$\rho_{M_tY_t} = bm + ab\rho_{X_tM_t} + my\rho_{M_tY_t} + ay\rho_{X_tY_t}. \quad \text{(A16)}$$

Rearranging terms and substituting from Equations A7 and A12 yields

$$\rho_{M_tY_t} = \frac{bm}{(1 - my)} + \frac{a^2bx}{(1 - mx)(1 - my)(1 - xy)}. \quad \text{(A17)}$$

One More Correlation

For our purposes, it is also useful to develop the expression for the stability of *M*: $\rho_{M_tM_{t-1}}$. Starting with the longitudinal model depicted in Figure 5, we see that

$$M_t = mM_{t-1} + aX_{t-1} + \varepsilon_{M_t}. \quad \text{(A18)}$$

The correlation of M_t with M_{t-1} can be expressed as

$$\rho_{M_tM_{t-1}} = m + a\rho_{X_tM_t}. \quad \text{(A19)}$$

Appendix B

Derivation of the Difference Between Cross-Sectional and Longitudinal Indirect Effects in Autoregressive Model

From the tracing rule (Kenny, 1979, pp. 30–34), the product of the paths a' and b' can be expressed as

$$a'b' = \rho_{X_tY_t} - c'. \quad \text{(B1)}$$

Appendix A shows that the correlation between *X* and *Y* is given by

$$\rho_{X_tY_t} = \frac{abx^2}{(1 - mx)(1 - xy)}. \quad \text{(B2)}$$

Substituting Equation B2 into Equation B1 yields

$$a'b' = \frac{abx^2}{(1 - mx)(1 - xy)} - c'. \quad \text{(B3)}$$

It then follows that the difference between $a'b'$ and ab can be written as

$$a'b' - ab = \frac{ab[x^2 - (1 - mx)(1 - xy)]}{(1 - mx)(1 - xy)} - c'. \quad \text{(B4)}$$

Appendix C

Derivation of Cross-Sectional Correlations Based on the Random Effects Model

In the case of complete mediation, Kenny et al.'s (2003) model stipulates that X, M, and Y for individual i at time t are related as follows:

$$M_{it} = d_{1i} + a_i X_{it} + e_{it}, \text{ and} \tag{C1}$$

$$Y_{it} = d_{2i} + b_i M_{it} + f_{it}. \tag{C2}$$

The following derivations assume that d_{1i}, d_{2i}, a_i, and b_i are all random effects. A further assumption is that the error terms (i.e., e_{it} and f_{it}) do not correlate with one another or with any of the random effects in the model. In addition, the distribution of all random effects and error terms is assumed to be multivariate normal.

Expression for the *XM* Correlation

From Equation C1, the covariance between X and M at a fixed time t can be written as

$$C(X_{it}, M_{it}) = C(X_{it}, d_{1i} + a_i X_{it} + e_{it}). \tag{C3}$$

Equation C3 can be rewritten as

$$C(X_{it}, M_{it}) = \sigma_{Xd_1} + C(X_{it}, a_i X_{it}). \tag{C4}$$

On the basis of Bohrnstedt and Goldberger (1969), the rightmost term of Equation C4 can be written as

$$C(X_{it}, a_i X_{it}) = a\sigma_X^2 + \mu_X \sigma_{aX} \tag{C5}$$

under multivariate normality. With an additional assumption that X and M are standardized (across individuals at time t), Equation C4 simplifies to

$$\rho_{XM} = a + \sigma_{Xd_1}. \tag{C6}$$

Expression for the *XY* Correlation

From Equation C2, the covariance between X and Y at a fixed time t can be written as

$$C(X_{it}, Y_{it}) = C(X_{it}, d_{2i} + b_i M_{it} + f_{it}). \tag{C7}$$

Equation C7 can be rewritten as

$$C(X_{it}, Y_{it}) = \sigma_{Xd_2} + C(X_{it}, b_i M_{it}). \tag{C8}$$

On the basis of Bohrnstedt and Goldberger (1969), the rightmost term of Equation C9 can be written as

$$C(X_{it}, b_i M_{it}) = b\sigma_{XM} + \mu_M \sigma_{bX} \tag{C9}$$

under multivariate normality. With an additional assumption that X and M are standardized (across individuals at time t), Equation C9 simplifies to

$$\rho_{XY} = \sigma_{Xd_2} + b\rho_{XM}. \tag{C10}$$

Because Equation C6 provides an expression for ρ_{XM}, Equation C10 can be rewritten as

$$\rho_{XY} = ab + \sigma_{Xd_2} + b\sigma_{Xd_1}. \tag{C11}$$

Expression for the *MY* Correlation

From Equation C2, the covariance between M and Y at a fixed time t can be written as

$$C(M_{it}, Y_{it}) = C(M_{it}, d_{2i} + b_i M_{it} + f_{it}). \tag{C12}$$

Equation C11 can be rewritten as

$$C(M_{it}, Y_{it}) = \sigma_{Md_2} + C(M_{it}, b_i M_{it}). \tag{C13}$$

On the basis of Bohrnstedt and Goldberger (1969), the rightmost term of Equation C13 can be written as

$$C(M_{it}, b_i M_{it}) = b\sigma_M^2 + \mu_M \sigma_{bM} \tag{C14}$$

under multivariate normality. With an additional assumption that X and M are standardized (across individuals at time t), Equation C14 simplifies to

$$\rho_{MY} = \sigma_{Md_2} + b. \tag{C15}$$

Equation C15 can be written in more basic terms by realizing that

$$C(M_{it}, d_{2i}) = C(d_{1i} + a_i X_{it} + e_{it}, d_{2i}). \tag{C16}$$

Equation C16 simplifies to

$$C(M_{it}, d_{2i}) = \sigma_{d_1 d_2} + a\sigma_{Xd_2}. \tag{C17}$$

Substituting Equation C17 into Equation C14 yields

$$\rho_{MY} = b + \sigma_{d_1 d_2} + a\sigma_{Xd_2}. \tag{C18}$$

(*Appendixes continue*)

Appendix D

More General Formulation of Random Effects Model

A more general formulation of the Kenny et al. (2003) mediation model would allow both M and Y to increase or decrease systematically with time irrespective of any influence on X. For example, M and Y might display straight-line growth even if X is neither systematically increasing nor decreasing. In such a case, a more general formulation of the model can be written by adding individual slopes to the model as depicted in Equations 18 and 19 of the text, resulting in models of the form

$$M_{it} = d^*_{1i} + d^*_{3i}t + a_iX_{it} + e_{it}, \text{ and} \quad \text{(D1)}$$

$$Y_{it} = d^*_{2i} + d^*_{4i}t + b_iM_{it} + f_{it}, \quad \text{(D2)}$$

where t represents the value of time.

The model shown in Equations 18 and 19 can be regarded as equivalent to the model shown in Equations D1 and D2 by conceptualizing d_{1i} and d_{2i} of Equations 18 and 19 as the status at time t of individual i on M controlling for X and on Y controlling for M, respectively. From this perspective, the derivations developed for this model in the text continue to apply, but the interpretation of the d_{1i} and d_{2i} is no longer the adjusted baseline but is instead the adjusted score at the specific moment in time t.

Alternatively, expressions could be derived in terms of the more general model shown in Equations D1 and D2. For example, consider the condition under which the bias in the cross-sectional direct effect will equal zero, as shown in Equation 25 of the text:

$$\sigma_{Xd_2} - (\sigma_{d_1d_2} + a\sigma_{Xd_2})(a + \sigma_{Xd_1}) = 0. \quad \text{(D3)}$$

This expression can be rewritten in terms of the straight-line growth model by realizing that

$$d_{1i} = d^*_{1i} + d^*_{3i}t, \text{ and} \quad \text{(D4)}$$

$$d_{2i} = d^*_{2i} + d^*_{4i}t. \quad \text{(D5)}$$

Substituting the expressions from Equations D4 and D5 into Equation D3 and rearranging terms shows that the condition for no bias in terms of the straight-line growth model parameters can be written as

$$\sigma_{Xd^*_2} + t\sigma_{Xd^*_4} - (\sigma_{d^*_1d^*_2} + t\sigma_{d^*_2d^*_3} + t\sigma_{d^*_1d^*_4} + t^2\sigma_{d^*_3d^*_4} + a\sigma_{Xd^*_2} + at\sigma_{Xd^*_4})(a + \sigma_{Xd^*_1} + t\sigma_{Xd^*_3}) = 0. \quad \text{(D6)}$$

Although Equation D6 looks even more daunting than Equation 25 in the text, they are in fact mathematically equivalent after taking into account the difference in meaning of the parameters. The important practical point is that regardless of which way the model is parameterized, very stringent conditions must hold for the cross-sectional analysis to yield an unbiased estimate of the longitudinal direct effect of X on Y. Although we have only demonstrated the correspondence between models for the direct effect, a similar correspondence also holds for the indirect effect.

Appendix E

Derivation of the Difference Between Cross-Sectional and Longitudinal Indirect Effects in Random Effects Model

As shown by Equation 26, the true average indirect effect in the random effects model is given by

$$E(a_ib_i) = ab + \sigma_{ab}. \quad \text{(E1)}$$

The bias in the cross-sectional estimate of the true random effects indirect effect can then be written as

$$bias = \beta_{MX}\beta_{YM\cdot X} - (ab + \sigma_{ab}). \quad \text{(E2)}$$

From Equations 15 through 17, $\beta_{YM\cdot X}$ can be expressed as

$$\beta_{YM\cdot X} = \frac{(b + \sigma_{d_1d_2} + a\sigma_{Xd_2}) - (a + \sigma_{Xd_1})(ab + \sigma_{Xd_2} + b\sigma_{Xd_1})}{1 - \rho^2_{XM}}. \quad \text{(E3)}$$

Substituting from Equations 15 and E3 into Equation E2 yields

$$bias = (a + \sigma_{Xd_1}) \times \left[\frac{(b + \sigma_{d_1d_2} + a\sigma_{Xd_2}) - (a + \sigma_{Xd_1})(ab + \sigma_{Xd_2} + b\sigma_{Xd_1})}{1 - \rho^2_{XM}}\right] - (ab + \sigma_{ab}). \quad \text{(E4)}$$

Received August 24, 2005
Revision received September 28, 2006
Accepted September 29, 2006 ■