Sensitivity Analysis of the No-Omitted Confounder Assumption in Latent Growth Curve Mediation Models

Davood Tofighi[1], **Yu-Yu Hsiao**[1], **Eric S. Kruger**[1], **David P. MacKinnon**[2], **M. Lee Van Horn**[1], and **Katie A. Witkiewitz**[1]

[1]University of New Mexico

[2]Arizona State University

Abstract

Latent growth curve mediation models are increasingly used to assess mechanisms of behavior change. For latent growth mediation model, like any another mediation model, even with random treatment assignment, a critical but untestable assumption for valid and unbiased estimates of the indirect effects is that there should be no omitted variable that confounds indirect effects. One way to address this untestable assumption is to conduct sensitivity analysis to assess whether the inference about an indirect effect would change under varying degrees of confounding bias. We developed a sensitivity analysis technique for a latent growth curve mediation model. We compute the biasing effect of confounding on point and confidence interval estimates of the indirect effects in a structural equation modeling framework. We illustrate sensitivity plots to visualize the effects of confounding on each indirect effect and present an empirical example to illustrate the application of the sensitivity analysis.

Keywords

Multiple mediation analysis; indirect effect; sensitivity analysis; latent growth curve; correlated augmented model

Mediation analysis assesses the effect of a treatment (independent) variable on an outcome variable via one or more mediators measured sequentially over time (MacKinnon, 2008). Complex multiple mediation models have become more common (e.g., Alvarez & Juang, 2010; Z. Chen & Wang, 2017; Demiray & Janssen, 2015; Kershaw, Mezuk, Abdou, Rafferty, & Jackson, 2010; Ma, Cheng, Ribbens, & Zhou, 2013; Priesemuth, Schminke, Ambrose, & Folger, 2014; Strelan, Karremans, & Krieg, 2017; Villegas-Gold & Yoo, 2014). In the current paper we focus on a mediation model with two covarying (parallel) mediators where the term "parallel" indicates that there is no direct causal path between the two mediators. One important application of a multiple mediator model with two covarying mediators is the latent growth curve mediation model (LGCMM) with repeated measures/multilevel data, whereby some treatment variable is expected to affect an outcome via the change in some mediating process (Selig & Preacher, 2009; von Soest & Hagtvet, 2011). LGCMM is

Correspondence concerning this article should be addressed to Davood Tofighi, Ph.D., Department of Psychology, 1 University of New Mexico, Albuquerque, New Mexico, 87131-0001. dtofighi@unm.edu.

increasingly used to test hypothesized mechanisms of behavior change following treatments for psychological disorders (Hallgren, Wilson, & Witkiewitz, 2018). For example, as shown in Figure 1, in alcohol treatment one might hypothesize that an anti-craving medication (e.g., naltrexone) affects reductions in alcohol use via changes in craving over time. Such a model would require estimation of two separate indirect effects: the indirect effect of treatment on alcohol use through the intercept of craving and the indirect effect of treatment on alcohol use through the slope of craving. In addition, LGCMM accounts for inter-dependencies of repeated measures data for the same individuals over time for the mediator and outcome variables[1].

Like any statistical model, a mediation model must meet certain assumptions to produce valid and unbiased estimates (James & Brett, 1984; Judd & Kenny, 1981; MacKinnon, 2008; Robins & Greenland, 1992). We focus on a widely acknowledged limitation of the mediation model, the assumption of no omitted confounders, which is necessary to produce unbiased, causally interpretable estimates of the indirect effect (Imai, Keele, & Tingley, 2010; Judd & Kenny, 1981; MacKinnon & Pirlott, 2015; Pearl, 2014; Robins & Greenland, 1992; Valente, Pelham, Smyth, & MacKinnon, 2017). When the independent variable is a random treatment assignment, this assumption states that there should be no omitted confounder variable that influences any pair of endogenous variables, that is, the mediator(s) and outcome(s). Unfortunately, this assumption is not testable (Holland, 1988). When designing a study, researchers usually include all the relevant variables that are hypothesized to influence the mediator(s) and the outcome variable(s). However, there are cases when the substantive theory is incomplete, it is unrealistic or impossible to measure all relevant variables, or a researcher analyzes an archival data set that does not include all theoretically relevant variables. Because it is typically impossible to randomize levels of the mediators to different participants, it is typically impossible to rule out the potential impact of omitted confounders on the relationships between the mediator(s) and outcome variables. It is thus recommended that one conduct a sensitivity analysis to ascertain the biasing effects of potential omitted confounders on the model estimates, including point and interval estimates of the indirect affect (Cox, Kisbu-Sakarya, Mio evi , & MacKinnon, 2013; Imai, Keele, & Yamamoto, 2010; MacKinnon & Pirlott, 2015; Tofighi & Kelley, 2016; VanderWeele, 2010).

For single level data (data that is not grouped or repeated measures), previous studies addressed sensitivity analysis for a) single mediator model, an example of which is depicted in Figure 2 (Cox et al., 2013; Imai, Keele, & Tingley, 2010; Valente et al., 2017; VanderWeele, 2010); b) multiple mediator model with two mediators that were assumed to be independent, which makes a strong causal assumption about independence of the two mediators (Imai & Yamamoto, 2013); and c) multiple mediator model with the mediators assumed to be sequentially related (Daniel, De Stavola, Cousens, & Vansteelandt, 2015; Harring, McNeish, & Hancock, 2017). To our knowledge, no study has extended sensitivity analysis to a longitudinal mediation model with two covarying mediators. The existence of two covarying mediators adds an additional complexity in computing indirect effects and

---

[1]Like latent growth curve model, linear mixed (multilevel) model is used to analyze multilevel data (Raudenbush & Bryk, 2002; Snijders & Bosker, 2011). For similarities and differences between the two techniques, please see these references (Curran, 2003; Willett & Bub, 2005).

conducting sensitivity analysis compared to a single mediator model. Analyzing each mediator one at a time would lead to biased estimates of indirect effects even when the two mediators are assumed to be independent (VanderWeele, 2015). A confounder may affect two relationships between each of the mediators and the outcome variable as opposed to one effect of the mediator on the outcome variable in a single mediator model. The effect of a potential confounder on the two relationships between the mediators and the outcome variable across two parallel mediation processes is likely to be differential, thus making modeling potential confounding bias more challenging. Thus, techniques developed for a single mediator model cannot directly be applied to a multiple mediator model.

For the LGCMM in Figure 1, we are interested in the two Between indirect effects through two covarying latent intercept and slope. None of the previous studies have extended sensitivity analysis to a longitudinal model with two covarying latent mediators. As mentioned before, applying single-mediator sensitivity techniques to a two-mediator model is likely to produce biased results and thus is not recommended (VanderWeele, 2015). Using multilevel structural equation modeling (SEM) framework, Tofighi and Kelly (2016) proposed a post-hoc method to compute the omitted confounder effect on the point estimate of the indirect effect but did not offer a method to compute standard errors and CIs for adjusted indirect effects. Using the potential outcomes framework (Rubin, 1974, 1978), Talloen et al.(2016) used fixed effect techniques (Allison, 2005) to remove the confounding bias at the Between level and thus relax the assumption of no omitted confounder at the Between level. The fixed effect technique cannot be applied to LGCMM in our study because the fixed-effect technique removes all the Between variability (McNeish & Kelley, 2018), which is the focus of our study. In related work, Bind et al. (2016) extended the potential outcomes framework to define indirect effects for a single mediator with longitudinal data; however, they did not propose a method to conduct sensitivity analysis. Moreover, techniques developed for single level data where the observations are assumed to be independent cannot directly be applied in a longitudinal growth context without additional analytic work to account for correlated observations (Tofighi & Kelley, 2016). Multilevel data poses additional challenges in determining whether mediation occurs at Between, Within, or cross levels (Bind et al., 2016; Tofighi & Kelley, 2016; Tofighi, West, & MacKinnon, 2013).

Our manuscript provides tools to help applied users conduct sensitivity analysis for an LGCMM with commonly used measures of effect sizes. We extend sensitivity analysis to LGCMM with two covarying latent growth factors in an SEM framework. We propose a technique termed *correlated augmented model sensitivity analysis* (CAMSA) to conduct sensitivity analysis. This technique, which is based on an SEM framework, uses a correlated augmented model. We proposed the correlated augmented model to prevent the negative residual variance problem of using latent proxy (phantom) variable techniques (Harring et al., 2017; Tofighi & Kelley, 2016), in which we reparametrized the model with a latent proxy variable technique when conducting sensitivity analysis. Reparameterization of an SEM has been suggested as a method to prevent negative residual variance, although not in the sensitivity analysis context (Rindskopf, 1983). The proposed CAMSA augments the mediation model with additional covariances between residuals associated with endogenous variables, termed confounder covariances, to model the effect of the potential confounders.

We offer analytic results to use confounder covariances to compute correlation between the residuals, termed sensitivity parameters (Imai, Keele, & Tingley, 2010). Because sensitivity parameters are confounder correlations, they offer an intuitive way to quantify the magnitude of confounding bias in terms of effect sizes. We also compute effect sizes for confounding effects in terms of the squared (semi-) partial correlations, which determine unique effect sizes due to the confounder. We present analytic results showing how the confounder correlations (sensitivity parameters) are used to estimate confounding biases, thereby modeling confounding effects. We also discuss strengths and weaknesses of CAMSA. We then use an empirical example to show how CAMSA can be conducted in Mplus (L. K. Muthén & Muthén, 2017). We also present two sets of graphs to visualize the results of the sensitivity analysis for each mediation process.

Causal Assumptions in Latent Growth Curve Mediation Modeling

In this section, we examine the causal inference assumptions necessary to produce an interpretable causal estimate of the indirect effect in an LGCMM. First, to clarify the no omitted confounder assumption, we consider a randomized experiment single mediation model in which X denotes a properly executed random assignment to two levels of a treatment. Because X denotes random assignment, we can rule out the following confounding effects: $C \rightarrow X$ and $X \rightarrow C$, given the relationships between X and the confounders (C) can be assumed to be zero because proper randomization balances out potential pre-treatment differences on the pre-treatment variables measured in C. Finally, when X is randomized, the $X \rightarrow M$ effect is not confounded by omitted variables (Holland, 1988). However, randomized assignment cannot rule out the following confounding effects. First, $X \rightarrow U$ effect, the confounder can be caused by X and cannot automatically be assumed zero because U is a post-treatment variable. Second, the effect of U on the $M \rightarrow Y$ relationship may not be ruled out. Thus, when the no-omitted confounder assumption is violated, the $M \rightarrow Y$ and $X \rightarrow Y$ effects in Figure 2 are potentially confounded by an unmeasured variable (Imai, Keele, & Tingley, 2010; Judd & Kenny, 1981; MacKinnon, 2008; Pearl, 2011, 2014; Robins & Greenland, 1992; VanderWeele, 2010).

Extension of the no-confounder assumption to a multilevel single mediator model has been discussed in the literature (Bind et al., 2016; Talloen et al., 2016; Tofighi & Kelley, 2016) We discuss extension of the no omitted confounder assumption to the LGCMM in Figure 1. Given that the hypothesized LGCMM has a randomized intervention (X), the no omitted confounder assumption is satisfied if the following two assumptions hold. 1- For each person, there should not be an omitted confounder that influences a pair of variables that include the two mediators as well as the outcome variable. 2- An omitted confounder should not be influenced by X. As mentioned previously, even for randomized intervention X, we cannot be certain the Assumption 1 holds, thus we should conduct sensitivity analysis. For the LGCMM model in the empirical example, however, we can assume that the randomized trial does not affect an omitted confounder that, in turn, affects the mediators and outcome variables. This is because in our substantive example, the intervention was specifically designed to manipulate alcohol craving. However, existence of a confounder U that is affected by X would imply that U is another mediator that is missing from the model. Given that X was randomized, and the treatment was specifically designed to reduce craving, we

can rule out that X does not impact any additional intermediate variable. Finally, given X is randomized, we can rule out the effect of an omitted confounder on the mediators. Taken together, multiple mediators (intercept and slope of the LGCMM) adds more complexity for conceptualizing potential confounders because there can be more patterns of confounding relationships compared to the single mediator model.

Proposed Sensitivity Analysis Method for LGCMM

In the next section, we specify two equivalent augmented models to compute the effect of omitted confounders. The analytic results will be used to conduct sensitivity analysis for a mediation model with two covarying mediators and one outcome variable. We use LGCMM to derive the results. Later, we discuss strengths and limitations of the results in terms of their application to different mediation models with two covarying mediators that are either latent or observed. The first model is termed the latent augmented model. This technique, which is based on SEM framework, is an extension of a sensitivity analysis method for a single mediator model in multilevel SEM (Tofighi & Kelley, 2016), as well as phantom variable sensitivity analysis in single level SEM (Harring et al., 2017). The second model is termed the correlated augmented model that uses correlations between residuals of the endogenous variables, termed confounder correlations (sensitivity parameters), to model the effect of the omitted confounder. As will be discussed later, the correlated augmented model has an advantage of not using a phantom variable, a latent variable without indicators, which could cause a negative residual variance.

Latent Augmented Model

In the latent augmented model, which is an extension of the sensitivity analysis proposed by Tofighi and Kelley (2016) and Harring et al. (2017), we first introduce a latent proxy (phantom) variable ϖ (a latent variable with no indicators), which represents all potential confounders into the model. This model is termed the latent augmented model. This latent proxy variable may represent one or more omitted confounders that are assumed to be linearly related to all endogenous variables in the model. Figure 3 shows the augmented LGCMM model. The dashed circle shows the latent proxy (phantom) variable ϖ and dashed arrows show unknown confounding effects on the mediators (latent intercept and slope) and the outcome variable. We term these coefficients as the confounder parameters. For identification purposes, the latent proxy variable is assumed to have a standard normal distribution (Tofighi & Kelley, 2016). The confounder parameters are set to values determining the degrees of confounding between the latent proxy variable and the endogenous variables (two latent variables and the outcome variable) in the model.

The following equations specify the latent augmented model. We use subscript "*" to denote the model parameters and residuals in the augmented model.

$$m_{ij} = \eta_{i0} + \eta_{i1} \, t_{ij} + e^*_{ij} \quad (1)$$

$$\eta_{i0} = \alpha_0^* + \gamma_{01}^* \, x_i + \gamma_{02}^* \, \varpi_i + \zeta_{0i}^* \quad (2)$$

$$\eta_{i1} = \alpha_1^* + \gamma_{11}^* \, x_i + \gamma_{12}^* \, \varpi_i + \zeta_{1i}^* \quad (3)$$

$$y_i = \nu^* + \kappa^* \, x_i + \lambda_1^* \eta_{i0} + \lambda_2^* \, \eta_{i1} + \lambda_3^* \, \varpi_i + \varepsilon_i^* \quad (4)$$

where the subscript i denotes a person $i = 1, \ldots, N$, subscript j denotes an occasion, $j = 1, \ldots, p$. The variables m_{ij} and t_{ij} denote the mediator (e.g., craving) and time score (e.g., week) for person i at occasion j, respectively; y_i and x_i denote the distal (ultimate) outcome (alcohol use) and random treatment assignment, respectively. The growth factors for the augmented LGCMM are denoted by η_{i0} for the intercept, describing person's latent score at the time score of 0 (e.g., mean craving in week 4), and η_{i1} for the latent slope, describing a person's latent rate of change in the mediator (e.g., weekly change in craving) over time; α_0^* and α_1^* denote the intercepts for the latent growth factors and ν denotes the intercept for the distal outcome; γ_{01}^*, γ_{11}^*, and κ^* quantify direct treatment effects on the latent intercept and slope and the distal outcome variable, respectively; λ_1^* and λ_2^* quantify the effect of the latent intercept and slope on the distal outcome variable, respectively. Confounder coefficients are the effects of ϖ on the latent intercept, slope, and the outcome variable, denoted by γ_{02}^*, γ_{12}^*, and λ_3^*, respectively. The confounder parameters, γ_{02}^*, γ_{12}^*, and λ_3^*, must be fixed to make the model identified for estimation purposes.

For the augmented model, the vector of residuals is denoted by $\boldsymbol{\varepsilon}^* = \begin{pmatrix} \mathbf{e}^*, & \varepsilon_i^*, & \zeta_{0i}^*, & \zeta_{1i}^* \end{pmatrix}^{\mathrm{T}}$, where $\mathbf{e}^* = \begin{pmatrix} e_{i1}^*, \ldots, e_{ip}^* \end{pmatrix}^{\mathrm{T}}$ is the vector of residuals for m_{ij}s, and T denotes vector transpose operator. It should be noted that ϖ is an exogenous latent variable without a residual term (i.e., residual term equals zero) and thus is not included in the residual vector. Given ϖ and $\boldsymbol{x}$, the vector of residuals has a multivariate normal distribution with mean vector of zeros, $\boldsymbol{\varepsilon}_i^* \sim \mathcal{N}(\mathbf{0}, \quad \boldsymbol{\Psi}^*)$ and the following covariance matrix:

$$\boldsymbol{\Psi}^* = \mathrm{cov}(\boldsymbol{\varepsilon}_i^* \mid \varpi, \mathbf{x}) = \begin{bmatrix} \boldsymbol{\Psi}_{\mathbf{e}^*} & 0 & 0 & 0 \\ 0 & \sigma_{\varepsilon^*}^2 & 0 & 0 \\ 0 & 0 & \sigma_{\zeta_0^*}^2 & \sigma_{\zeta_0^*, \zeta_1^*} \\ 0 & 0 & \sigma_{\zeta_0^*, \zeta_1^*} & \sigma_{\zeta_1^*}^2 \end{bmatrix}$$

where $\boldsymbol{\Psi}_{\mathbf{e}^*}$ is the covariance matrix of $\mathbf{e}^*$. When all confounders are included in the model, it can be shown that the residuals associated with mediators and outcome variables are uncorrelated under certain additional assumptions (Tofighi & Kelley, 2016; Tofighi et al., 2013). This means that residuals associated with the outcome variable ε_i^* are not correlated with the residuals associated with latent factors, ζ^*s. It should be noted that, however, ζ^*s, are usually assumed to covary in multilevel and longitudinal analysis literature (Grimm, Ram, & Estabrook, 2017; Singer & Willett, 2003; Snijders & Bosker, 2011). This covariance can be of substantive interest because the status for an individual at time zero (intercept) can be related to the person's rate of growth (slope). However, we should emphasize that in the augmented model this covariance may not be solely due to an omitted confounder. In other words, if this covariance is a result of an omitted confounder, then this covariance is already modeled by ϖ, which is assumed to represent all the omitted confounders. As shown below, given $\boldsymbol{x}$ the covariance between the growth factors is decomposed into two parts:

$$\text{cov}\big(\eta_{i0}, \eta_{i1} \,\big|\, \boldsymbol{x}\big) = \gamma_{02}^* \gamma_{12}^* \sigma_\varpi^2 + \sigma_{\zeta_0^*, \zeta_1^*}, \quad (5)$$

where $\sigma_{\zeta_0^*, \zeta_1^*} = \text{cov}(\zeta_{i1}^*, \zeta_{i2}^* | x)$. In the above equation, the first term $\sigma_\varpi^2 \; \gamma_{02}^* \gamma_{12}^*$ quantifies the confounding part of the covariance between the growth factors due to ϖ and the term $\sigma_{\zeta_0^*, \zeta_1^*}$ specifies the non-confounding part of the covariance. Similarly, we can derive the covariance between η_{i0} and η_{i1} and ϖ:

$$\text{cov}\big(\eta_{i0} \,,\, \varpi_i \,\big|\, \boldsymbol{x}\big) = \gamma_{02}^* \, \sigma_\varpi^2, \quad (6)$$

$$\text{cov}\big(\eta_{i1} \,,\, \varpi_i \,\big|\, \boldsymbol{x}\big) = \gamma_{12}^* \, \sigma_\varpi^2, \quad (7)$$

More succinctly, the conditional covariance matrix between the vector latent variables, $\boldsymbol{\eta}_i^* = \big(\eta_{01}, \quad \eta_{i1}, \quad \varpi_i\big)^{\mathrm{T}}$, is as follows:

$$\text{cov}(\boldsymbol{\eta}_i^* | \boldsymbol{x}) = \sigma_\varpi^2 \begin{bmatrix} (\gamma_{02}^*)^2 + \sigma_{\zeta_0^*}^2 & \gamma_{02}^* \gamma_{12}^* + \sigma_{\zeta_0^*, \zeta_1^*} & \gamma_{02}^* \\ \gamma_{02}^* \gamma_{12}^* + \sigma_{\zeta_0^*, \zeta_1^*} & (\gamma_{12}^*)^2 + \sigma_{\zeta_1^*}^2 & \gamma_{12}^* \\ \gamma_{02}^* & \gamma_{12}^* & 1 \end{bmatrix} \quad (8)$$

To compute the indirect effect in the augmented model, we introduce different levels of confounding bias by manipulating values of the confounding parameters. The confounding parameters take on various values influencing the latent intercept and slope and the outcome

variable. We need to identify different values for the confounder parameters that are of substantive interest for two reasons. First, note that because the mediators and outcome variables are not standardized, the parameter values depend on the metric of the variables. Second, we need computational formulas that would translate a researcher's knowledge about the range of correlation values to confounder parameters. To make the choice of the values for the parameters more intuitive, we compute the confounder parameter values as a function of the correlation values which are more understandable. The correlation values, termed confounder correlations (sensitivity parameters), can be used to compute effect sizes of confounding effects in terms of the squared (semi-)partial correlations, which determine unique effect sizes due to the confounder. That is, we first determine a set of plausible values (e.g., correlation values based on prior research if any exists or other information) for confounder correlations and then compute the confounder parameter values. Given that ϖ and x are not correlated and variance of ϖ is one, we have

$$\gamma_{02}^* = \sigma_{\eta_0} \, \rho_{\eta_0, \varpi} \quad (9)$$

$$\gamma_{12}^* = \sigma_{\eta_1} \quad \rho_{\eta_1, \varpi} \quad (10)$$

$$\lambda_3^* = \sigma_y \, \rho_{y, \varpi} - \lambda_1^* \, \gamma_{02}^* - \lambda_2^* \, \gamma_{12}^* \quad (11)$$

where $\rho_{\eta_0, \varpi}$, $\rho_{\eta_1, \varpi}$, and $\rho_{y, \varpi}$ denote confounder correlations between the latent proxy variable and latent intercept and slope and the outcome variable, respectively. These correlations take on a combination of the values set by the researcher. The parameters, σ_{η_0}, σ_{η_1}, and σ_y are the population standard deviations for the latent intercept, latent slope and outcome variable, respectively. We can estimate the variance of the growth factors ($\sigma_{\eta_0}^2$ and $\sigma_{\eta_1}^2$) from the latent growth model where there are no predictors (i.e., x and C) or the outcome variable (y) in the model; σ_y is estimated by calculating the sample standard deviation.

Unfortunately, the latent augmented model (as specified in Figure 3) can result in a negative residual variance (i.e., Heywood cases; Dillon, Kumar, & Mulani, 1987; Kolenikov & Bollen, 2012; Rindskopf, 1984). For our empirical example, the latent augmented LGCMM produced a negative variance estimate regardless of the values of sensitivity parameters. The Heywood case for an augmented LGCMM is likely to occur because fitting a phantom variable that is not supported by data is likely to cause negative residual variance. As Bartholomew, Knott, and Moustaki points out (2011),

> " One of the commonest cause of Heywood cases is the attempt to extract more factors than are present. This is readily demonstrated by simulation but might have been anticipated on the grounds that artificially inflating the communality forces the residuals towards zero." (p. 67)

Correlated Augmented Model Sensitivity Analysis

To prevent the Heywood case associated with estimation of the latent augmented model, we propose the *correlated augmented model sensitivity analysis* (CAMSA). CAMSA uses the correlated augmented model[2] shown in Figure 4, which is equivalent (i.e., has the same likelihood function; Kreft, de Leeuw, & Aiken, 1995) to the latent augmented model in Figure 3. That is, we reparametrize the latent augmented model to create the correlated augmented model that does not use a latent proxy (phantom) variable and thus avoids a potential Heywood case (i.e., negative residual variance). Although not in the context of sensitivity analysis, previous research suggested using reparameterization to prevent Heywood cases (Rindskopf, 1983). Because the two models, the reparametrized model and the original model, are equivalent, the reparametrized model parameters are a function of the parameters in the original model. The corresponding parameters in the equivalent models are not necessarily equal, but there must be a one-to-one transformation between the corresponding parameters in the equivalent models (Kreft et al., 1995). As a result, we can use the parameter estimates of the reparametrized model to compute the parameter estimates of interest in the original model and vice versa.

We next present the equations for the correlated augmented model shown in Figure 4. We use superscript "**" to denote the parameters in the correlated augmented model.

$$m_{ij} = \eta_{0i} + \eta_{1i}\, t_j + e_{ij}^{**} \quad (12)$$

$$\eta_{0i} = \alpha_0^{**} + \gamma_{01}^{**}\, x_i + \zeta_{0i}^{**} \quad (13)$$

$$\eta_{1i} = \alpha_1^{**} + \gamma_{11}^{**}\, x_i + \zeta_{1i}^{**} \quad (14)$$

$$y_i = \nu^{**} + \kappa^{**}\, x_i + \lambda_1^{**}\, \eta_{0i} + \lambda_2^{**}\, \eta_{1i} + \varepsilon_y^{**} \quad (15)$$

[2]Correlating residuals as a place holder to model an effect has been suggested in SEM literature. For example, Cole, Ciesla, and Steiger (2007) argue for using correlated residuals in a latent variable model to estimate common method (common method variance) effect. We used the same concept to account for the omitted confounder effect in CAMSA except that we fix the confounder covariances between the residuals to specific values.

Similar to the latent augmented model in (1)–(4), m_{ij} and t_{ij} denote the mediator and time score for person i at occasion j, respectively; y_i and x_i denote the distal outcome and random treatment assignment, respectively. In the equations for the latent growth factors η_{i0} and η_{10}, a^{**}s and ζ^{**}s are the intercepts and residuals while γ^{**}s are the treatment effects on the growth factors. In the equation for y_i, κ^{**}, λ_1^{**}, and λ_2^{**} quantify the partial regression effects of x_i, η_{i0} and η_{10} on y_i, respectively; ν^{**} and ε_y^{**} are the intercept and residual, respectively.

It is important to note that there are subtle differences in interpreting the covariances (correlations) between the residuals in the correlated augmented model. The two covariances between the residuals associated with latent intercept and slope and the outcome variable, $\sigma_{\varepsilon_y^{**},\zeta_0^{**}}$ and $\sigma_{\varepsilon_y^{**},\zeta_1^{**}}$, respectively, model the confounding effects of the omitted confounders, and thus are termed *confounder covariances*. We show analytically that the following relationships between the confounder covariances in the correlated augmented model and the confounder parameters in the latent augmented model hold:

$$\sigma_{\varepsilon_y^{**},\zeta_0^{**}} = \gamma_{02}^* \, \lambda_3^* \, \sigma_\varpi^2, \quad (16)$$

$$\sigma_{\varepsilon_y^{**},\zeta_1^{**}} = \gamma_{12}^* \, \lambda_3^* \, \sigma_\varpi^2, \quad (17)$$

$$\rho_{\varepsilon_y^{**},\zeta_0^{**}} = \frac{\gamma_{02}^* \, \lambda_3^* \, \sigma_\varpi^2}{\sqrt{(\lambda_3^*)^2 \, \sigma_\varpi^2 + \sigma_{\varepsilon_y^*}^2} \, \sqrt{(\gamma_{02}^*)^2 \, \sigma_\varpi^2 + \sigma_{\zeta_0^*}^2}}, \quad (18)$$

$$\rho_{\varepsilon_y^{**},\zeta_1^{**}} = \frac{\gamma_{12}^* \, \lambda_3^* \, \sigma_\varpi^2}{\sqrt{(\lambda_3^*)^2 \, \sigma_\varpi^2 + \sigma_{\varepsilon_y^*}^2} \, \sqrt{(\gamma_{12}^*)^2 \, \sigma_\varpi^2 + \sigma_{\zeta_1^*}^2}}, \quad (19)$$

where the confounder parameters $\gamma_{02}^*, \gamma_{12}^*$, and λ_3^* denote the effect of latent proxy variable in (2)–(4) for the latent augmented model; term σ_ϖ^2 denotes the latent proxy variance, which is assumed to be one. The product terms $\gamma_{02}^* \lambda_3^*$ and $\gamma_{12}^* \lambda_3^*$ quantify the confounding effects of the omitting the latent proxy variable ϖ on the covariances (correlations) between latent intercept and outcome variable and between the latent slope and outcome variable, respectively. In summary, in the correlated augmented model, the $\sigma_{\varepsilon_y^{**},\zeta_0^{**}}$ and $\sigma_{\varepsilon_y^{**},\zeta_1^{**}}$ are confounder covariances quantifying the effects of omitted confounders, which, in the

latent augmented model, equal to the confounding effects of the latent proxy variable on two relationships: latent intercept to outcome variable and latent slope to outcome variable.

The residual covariance between the growth factor residuals, $\sigma_{\zeta_0^{**}, \zeta_1^{**}}$, in the correlated augmented model has a different interpretation from the corresponding covariance, $\sigma_{\zeta_0^{*}, \zeta_1^{*}}$, in the latent augmented model. As shown in Equation (5) for the latent augmented model, the term $\sigma_{\zeta_0^{*}, \zeta_1^{*}}$ quantifies the covariance between the growth factor residuals that is *not* due to the latent proxy variable, and thus is not due to the omitted confounders. However, as shown below, in the correlated augmented model, the covariance between the growth factor residuals is the sum of two terms that capture both confounding and non-confounding effects:

$$\sigma_{\zeta_0^{**}, \zeta_1^{**}} = \sigma_\varpi^2 \gamma_{02}^* \gamma_{12}^* + \sigma_{\zeta_0^{*}, \zeta_1^{*}}. \quad (20)$$

The first part in (20), the term $\sigma_\varpi^2 \gamma_{02}^* \gamma_{12}^*$ quantifies the confounding covariance that is modeled by the latent proxy variable in the latent augmented model; the confounder parameters γ_{02}^* and γ_{12}^* quantify the effects of the latent proxy variable on the growth factors as shown in (2) and (3). The second part in (20), the term $\sigma_{\zeta_0^{*}, \zeta_1^{*}}$ is the covariance between the growth factor residuals that is not due to the confounders.

Problematic Values of Confounder Correlation

One issue concerning estimating both the latent augmented model and correlated augmented model is *empirical under-identification*, which would result in unstable parameter estimates with extremely large standard errors (Kenny & Milan, 2012; Rindskopf, 1984). As Kenny (2004, p. 51) noted, "[e]mpirical under-identification is defined by zero or near-zero denominators in the estimates of structural parameters". Empirical under-identification would likely result in a negative residual variance (F. Chen, Bollen, Paxton, Curran, & Kirby, 2001), which can occur with CAMSA and latent augmented model. In addition, it might problematic when certain combination of the confounder correlation values is selected without any restriction, and thus one or more correlation values can take on non-admissible (out of bound) values. For instance, for correlations between X, M, and Y, each correlation is bounded as follows (Leung & Lam, 1975):

$$\rho_{MX}\rho_{XY} - \sqrt{\left(1 - \rho_{MX}^2\right)\left(1 - \rho_{XY}^2\right)} \le \rho_{MY} \le \rho_{MX}\rho_{XY} + \sqrt{\left(1 - \rho_{MX}^2\right)\left(1 - \rho_{XY}^2\right)}$$

When encountering four or more variables, as is the case for a two-mediator model such as LGCMM, restrictions between the correlation values become more complex. A correlation matrix has a special property that defines the restrictions between correlations: a correlation matrix must be positive semi-definite (PSD) (Rousseeuw & Molenberghs, 1994). A matrix is said to be PSD if and only if its determinant is greater than or equal to zero. Not all

combinations of the confounder correlation values picked by a researcher would result in PSD correlation matrices implied by the model. One advantage of CAMSA as well as the latent augmented model is that SEM software, such as Mplus (L. K. Muthén & Muthén, 2017), would generate an error for inadmissible confounder correlation values that would result in a non-PDS covariance matrix. The corresponding confounder correlation values can then be excluded from the sensitivity analysis.

Application of Sensitivity Analysis to Motivating Example

In this section, we first estimate LGCMM using an empirical example. Next, we describe conducting CAMSA for the empirical example.

Empirical Example

We used data collected from the COMBINE study ["Combined Pharmacotherapies and Behavioral Interventions for Alcohol Dependence" (The COMBINE Study Research Group, 2003)]. In the COMBINE trial, eligible participants were randomly assigned to one of nine treatment conditions that included a combination of prescription medications (acamprosate, naltrexone, or placebo equivalents) and behavioral interventions (medication management [MM] or combined behavioral intervention [CBI]): 1) MM + naltrexone, 2) MM + acamprosate, 3) MM + naltrexone + acamprosate, and 4) MM + placebo. Five additional conditions included CBI: 5) CBI + naltrexone, 6) CBI+ acamprosate, 7) CBI+ naltrexone + acamprosate, 8) CBI+ placebo, and 9) CBI only. For this study, we were interested in the effect of the treatment conditions that included active naltrexone verses the treatment conditions that administered placebo naltrexone.

The goal of the empirical example was to evaluate the effect of naltrexone on alcohol use (outcome) via changes in craving (mediator). Self-reported craving was assessed by the Penn Alcohol Craving Scale (PACS; Cronbach's α .85). The PACS consists of 5 items that assess frequency, intensity and duration of alcohol craving and an overall rating of craving for the prior week (Flannery, Volpicelli, & Pettinati, 1999). For this study, we used PACS craving scores measured in weeks 1, 2, 4, 6, 8, 10, 12. The outcome variable was the percent drinking days of any alcohol during the prior month at week 16, assessed by the Form 90 (Miller, 1996). To increase numerical stability during estimation, the percent drinking days was divided by 10 to make its dispersion smaller and more comparable to the rest of the data.

We estimated the LGCMM in Mplus 8.0 (L. K. Muthén & Muthén, 2017). As seen in Figure 1, naltrexone significantly reduced alcohol craving B(SE) = −1.154 (0.362), p =.001, 95% CI [−1.864; −0.445], but did not influence the linear change in alcohol craving over time B (SE) =0.004 (0.039), p = .915, 95% CI: [−0.071; 0.080]. Controlling for the naltrexone treatment, the effect of the mean alcohol craving at week 4 and the weekly growth rate of alcohol craving on percentage drinking days were both significant, B (SE) = 0.197 (0.015), p <.001, 95% CI: [0.167, 0.226] and B (SE) =2.451 (0.205), p <.001, 95% CI: [2.048, 2.853], respectively. Assuming the correct specification assumptions, including the no-omitted-confounder assumption holds, we use Bonferroni adjusted $\alpha/2$ = .025 and the distribution of the product of coefficients approach to compute a CI (Tofighi & MacKinnon, 2011), for the

indirect effect involving latent intercept, B (SE) = −0.227 (0.073), 97.5% CI [−0.398 −0.067]; for the indirect effect involving latent slope, B (SE) = 0.01 (0.094), 97.5% CI [−0.206 0.227]. Thus, naltrexone was effective in reducing percent drinking days in the last month of treatment (assessed at week 16) via the reduction in the mean craving observed during the first month of treatment (intercept; week 4), but there was no indirect effect of naltrexone on percent drinking days via change in craving over time.

Correlated Augmented Model Sensitivity Analysis (CAMSA)

We conduct CAMSA for the empirical example to assess if the indirect effects are robust to varying degrees of omitted confounder bias because the values of the mediators, latent intercept and slope, are not randomized. Thus, there could be other confounders that would influence the latent intercept and slope as well as the outcome variable. For example, genotype (Kranzler, Armeli, Covault, & Tennen, 2013), type of drinker (i.e., reward verses relief drinker; Mann et al., 2018), or smoking status (Schacht et al., 2017) could impact the effect of naltrexone on craving intercept/slope, as well as drinking outcomes. Code scripts and accompanying instructions to run the sensitivity analysis are available in the supplemental materials.

To calculate the values of the coefficients in the correlated augmented model, we first choose a set of plausible values for confounder correlation parameters, $\rho_{\varepsilon_y^* *, \zeta_0^* *}$ and $\rho_{\varepsilon_y^* *, \zeta_1^* *}$.

For each set of values, the correlated augmented model is estimated, and adjusted values of model parameters as well as the indirect effect and its 97.5% CI are calculated. In our example, we choose the values for the confounder correlations within the range of −.5 and .5, which also covers the correlation values ± .1, ± .3, and ± .5, corresponding to Cohen's (1988) guideline on small, medium, and large effect sizes. One may choose a different set of values based on prior research and/or knowledge of the magnitude and direction of potential omitted confounders.

Results

CAMSA produces point and interval estimates of the coefficients showing potential impact of varying degrees of confounding on all model parameters. We focus on the results of the sensitivity analysis on two indirect effect estimates and respective 97.5% CIs. As mentioned before, for each admissible set of confounder correlations values, sensitivity analysis would compute two indirect effects estimates and CIs in the correlated augmented model, resulting in hundreds of point and interval estimates. Non-admissible correlations are identified through the errors (e.g., "negative psi matrix") produced by Mplus and thus are excluded.

We present two sets of graphs to visually summarize varying degrees of confounding effects on the point and interval estimates for each indirect effect: sensitivity contour plot and sensitivity confidence band plot. The sensitivity contour plots, shown in Figure 5, display the magnitude of the indirect effect as a function of the confounder correlations. For example, panel-a in Figure 5 shows the contour plot for the indirect effect estimate through the mediator latent intercept. The *x*-axis and *y*-axis represent the range of values for the confounder correlations between the latent intercept and $Y(\rho_1)$ and latent slope and $Y(\rho_2)$, respectively. Each contour line illustrates all the combinations of the values for the two

confounder correlations that result in the same indirect effect estimate. A value for the indirect effect corresponding to each contour line is displayed adjacent to the line. For the indirect effect through the intercept, the estimates range roughly from −0.5 to −0.1 for ρ_1 . 2 and ρ_2 .5. Given the original estimate for the indirect effect is B (SE) = −0.227 (0.073), 97.5% CI [−0.398, −0.067], the point estimate of the adjusted indirect effect does not appear to change dramatically for the range of confounder correlations. However, as $\rho_2 > .2$, the sign of the indirect effect estimates changes. For indirect effect estimate through the slope, panel-b in Figure 5 shows that indirect effect estimate roughly ranges from −0.005 to 0.025 for −.5 ρ_1 and ρ_2 .5. Given the original estimate of the indirect effect is B (SE) = 0.01 (0.094), 97.5% CI [−0.206, 0.227], the indirect effect estimate did not fall outside of the original CI. In sum, it appears that the values of confounder correlations have considerable impact on the indirect effect through intercept in that sign of the indirect effect changes; on the other hand, the values of confounder correlation appear to have a minimal impact on the indirect effect through slope.

A second way to summarize the effect of omitted confounders is to present sensitivity confidence band plots to examine changes in the CIs. This is advantageous because one can examine the effect of the omitted confounders on the uncertainty about the indirect effects as well as the point estimate. Because there are many combinations of the confounder correlation we use the information from sensitivity contour plots to focus on the range of the confounder correlation values for which the inference about the indirect effect changes. That is, we look at the ranges where the status of a CI containing zero changes to not containing zero and vice versa. Figure 6 shows four panels of 97.5% confidence bands for the indirect effect through intercept; each panel corresponds to a fixed value of ρ_2 where the *x*-axis shows a more restricted range of values, 0 ρ_1 .5, and *y*-axis shows indirect effect values. We chose positive values for ρ_2 because the results were identical to the corresponding negative values. For example, the confidence band plots were identical for $\rho_2 = \pm 1$. Of course, one may choose different positive or negative values for ρ_2 in a different context. As the value of ρ_2 increases from small (.1) to medium (.3) and then to large (.5), the regions of non-significance corresponding to ρ_1 were roughly (.35, .40), (.30, .42), and (.25, .37), respectively. Below the lower limit of ρ_1 for each non-significance region, the CIs contain negative indirect effect estimates while above the upper limit of ρ_1 for the non-significance region, the CIs contain the positive indirect effect estimates. Figure 7 illustrates the panel plots of confidence bands for the indirect effect estimate through the slope for four values of ρ_1. The *x*-axis shows the values for ρ_2 between 0 and .5. In each panel, as ρ_2 becomes larger the confidence band becomes narrower and then widens. For all values of ρ_2, however, all the confidence bands contain zero. This shows the inference about the indirect effect through the slope is robust in that the CIs contain zero as the confounder correlation ρ_2 increases from zero to large.

Conclusion and Summary

Latent growth curve mediation models (LGCMM) have become more common in applied contexts to study whether some treatment variable is expected to affect an outcome via changes in some mediating process. Given the growing popularity of this model in the field (see Hallgren et al., 2018), it is critically important to check underlying assumptions to ensure valid results. One of these assumptions is an untestable assumption that there are not omitted confounders from the model. For LGCMM, the no omitted confounder assumption implies that there should be no omitted variable that influences the two mediators as well as the outcome variable. Furthermore, there should be no omitted variable that is influenced by the independent variable X that in turn would influence the mediator and outcome variables.

The current study provides an extension of other recent work (e.g., Bind et al., 2016; Harring et al., 2017; Imai et al., 2010; Tofighi & Kelley, 2016), by developing and testing a method for sensitivity analysis in LGCMM. We proposed a technique termed correlated augmented model sensitivity analysis (CAMSA) for models with two mediators, with an application to LGCMM. CAMSA uses a correlated augmented model to represent cumulative and linear effects of one or more potential confounders. We presented analytic results showing that the confounder covariances between residuals associated with the mediators (latent intercept and slope) and the outcome variable account for confounding effects. We presented formulas to compute the model parameters from the confounder correlations (sensitivity parameters) as the correlation values are more convenient to quantify confounding bias. For multiple combinations of the values for the confounder correlations, CAMSA estimates the correlated augmented model and generates point and interval estimates for the indirect effects, iteratively.

One advantage of the correlated augmented model is using confounder correlations (sensitivity parameters) to quantify confounder bias. Using confounder correlation is more convenient in that one can directly quantify the confounder effects in terms of confounder correlations given analytic formulas. In addition, the correlated augmented model is suitable for mediation models where the endogenous variables are latent. A second advantage of CAMSA is that it can be implemented within the SEM framework and estimated using available software, such as Mplus (L. K. Muthén & Muthén, 2017). We provided an application of CAMSA using an empirical example. An R (R Development Core Team, 2017) script to conduct sensitivity analysis using Mplus along with the instruction on how to use R and Mplus to conduct sensitivity analysis is provided in the Supplemental Materials.

Another noteworthy aspect of conducting sensitivity analysis is the summary and interpretation of point and interval estimates as a function of many combinations of the confounder correlation values. We presented sensitivity contour plots that visually illustrate point estimates of the two indirect effects as a function of confounder correlations, as well as a second visual summary of the results that is the sensitivity confidence band plot. For each indirect effect, the sensitivity confidence band plot illustrates the effect of confounder bias on the confidence intervals of the indirect effect estimates. We recommend researchers try several ranges of the values for both confounder correlations shown on the x-axis and the

panel to capture important features of the effects confounding on the interval estimates for an indirect effect.

As mentioned previously, the no-omitted confounder assumption is one of the several specification assumptions underlying a mediation model. Additional assumptions include correct functional form of the relationships between the variables, correct distributional assumptions of the residuals, no outliers, no non-random missing data due to attrition, and no measurement error (Baraldi & Enders, 2010; Fritz, Kenny, & MacKinnon, 2016; Fritz & MacKinnon, 2007; James & Brett, 1984; Judd & Kenny, 1981; MacKinnon, 2008). In addition, within the SEM framework, there are additional statistical distributional assumptions and sample size requirements that need to be satisfied when SEM is used to estimate a mediation model (Hoyle, 2012; Kline, 2016; West, Finch, & Curran, 1995). One must be prudent in examining the combined effect of violation of one or more of these assumptions along with violation of the no omitted confounder assumption. For example, a study showed that combined measurement error and omitted confounder effect can have either a heightened or mitigated combined biasing effect on the indirect effect (Fritz et al., 2016). It is recommended that one would account for the effect of the measurement error by using latent variables.

We also note CAMSA is not suitable for a mediation model with observed endogenous variables because the model would not be identified even though correlations between the residuals are fixed (B. Muthén & Asparouhov, 2015). For such models, we recommend one conduct sensitivity analysis using the latent (phantom variable) augmented model (Harring et al., 2017; Tofighi & Kelley, 2016). It is also possible to extend CAMSA to mediation models with three or more covarying as well as sequential related latent mediators. However, the extension requires additional consideration about confounder correlations and analytic derivation to model confounding effects on all the mediators and outcome variable. The extension remains a topic of future study. Much prior research on sensitivity analysis has focused on the randomized case. With nonrandomized X, there are confounders of the X to M relation. Consequently, equations for calculating sensitivity to confounding would be applied to the X and M relation, and then these equations would need to be developed to include sensitivity to confounding of both the X to M and M to Y relations. The analytical techniques in this paper could be extended to this case and represent a topic for future development.

One should be prudent in interpreting the results of the sensitivity analysis. Statistical sensitivity analysis results should be further justified by the researcher as to why the conclusion is consistent with substantive theory and previous research. When the results about the significance of the indirect effects change qualitatively based on values of confounder correlations then a researcher might need to further probe the model and its underlying theory, design of the study, and implementation of the proposed study, as well as check for additional statistical specification assumptions of the mediation model (MacKinnon, 2008; MacKinnon, Fairchild, & Fritz, 2007; MacKinnon & Pirlott, 2015; Preacher, 2015).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments.

Appendix

Analytic Results for the Latent Augmented Model

In deriving the above expression, we use the following relations.

a. $\text{cov}(\varpi, x_i) = 0$ because x_i is randomized, it neither influences nor is influenced by ϖ. That is, we assume ϖ and x_i to be orthogonal (independent)

b. $\text{cov}(\varpi, \zeta^*_{0i}) = \text{cov}(\varpi, \zeta^*_{1i}) = 0$ and $\text{cov}(\varpi, \zeta^*_{0i}|x) = \text{cov}(\varpi, \zeta^*_{1i}|x) = 0$. These expressions hold because we assume that ϖ models all the omitted confounders that influence η_0 and η_1, and ϖ and x_i are independent.

c. $\text{cov}(x_i, \zeta^*_{0i}) = \text{cov}(x_i, \zeta^*_{1i}) = 0$ and $\text{cov}(x_i, \zeta^*_{0i}|\varpi) = \text{cov}(x_i, \zeta^*_{1i}|\varpi) = 0$. These expressions hold because we assume that ϖ models all the omitted confounders that influence η_0 and η_1, and ϖ and x_i are independent.

d. We also assume constant variance: $\text{var}(\zeta^*_{0i}| \ x_i, \varpi) = \sigma^2_{\zeta^*_0}$, $\text{var}(\zeta^*_{1i}| \ x_i, \varpi) = \sigma^2_{\zeta^*_1}$, and $\text{var}(\varepsilon^*_i| \ x_i, \varpi, \ \eta_{i0}, \eta_{10}) = \sigma^2_{\varepsilon^*}$.

Now, we derive the following expressions. The expression for $\text{cov}(\eta_{i0}, \eta_{i1}| \ \boldsymbol{x})$ in (5) is derived as follows:

$$\text{cov}(\eta_{i0}, \eta_{i1} | \boldsymbol{x}) = \text{cov}(\ \alpha^*_0 + \gamma^*_{01} \ x_i + \gamma^*_{02}\varpi + \ \zeta^*_{0i}, \ \alpha^*_1 + \gamma^*_{11} \ x_i + \ \gamma^*_{12}\varpi + \ \zeta^*_{1i} \ |\boldsymbol{x})$$

$$= \text{cov}(\ \gamma^*_{02}\varpi + \ \zeta^*_{0i}, \ \gamma^*_{12}\varpi + \ \zeta^*_{1i} \ |\boldsymbol{x}) = \alpha^2_\varpi \gamma^*_{02}\gamma^*_{12} + \sigma_{\zeta^*_0, \zeta^*_1}$$

The expressions in (6) and (7) are obtained as follows:

$$\sigma_{\eta_0, \varpi|x} = \text{cov}(\alpha^*_0 + \gamma^*_{01} \ x_i + \gamma^*_{02}\varpi + \ \zeta^*_{0i}, \varpi \ | \ \boldsymbol{x}) = \ \text{cov}(\gamma^*_{02}\varpi + \ \zeta^*_{0i} \ , \varpi \Big| \boldsymbol{x}) \quad \text{(A1)}$$

$$= \ \gamma^*_{02} \ \sigma^2_\varpi$$

$$\sigma_{\eta_1, \varpi | x} = \mathrm{cov}(\eta_1, \varpi \mid x) = \mathrm{cov}(\alpha_1^* + \gamma_{11}^* x_i + \gamma_{12}^* \varpi + \zeta_{1i}^*, \varpi \mid x) \quad \text{(A2)}$$
$$= \mathrm{cov}(\gamma_{12}^* \varpi + \zeta_{1i}^*, \varpi | x) = \gamma_{12}^* \sigma_\varpi^2$$

To derive the expressions in (9)–(11), we first need to obtain expressions for $\sigma_{\eta_0, \varpi | x}$ and $\sigma_{\eta_1, \varpi | x}$, given in (A1) and (A2), $\sigma_{\eta_0, \varpi}, \sigma_{\eta_1, \varpi}, \sigma_{\eta_0}^2 \sigma_{\eta_0 | x}^2, \sigma_{\eta_1}^2, \sigma_{\eta_1 | x}^2, \rho_{\eta_0, \varpi}, \rho_{\eta_0, \varpi | x}, \rho_{\eta_1, \varpi},$ $\rho_{\eta_1, \varpi | x}, \sigma_{y, \varpi}$ and $\rho_{y, \varpi}$. We first obtain the conditional and unconditional variance of the growth factors. Given (2), the unconditional variance of the latent intercept is as follows:

$$\sigma_{\eta_0}^2 = \mathrm{var}(\alpha_0^* + \gamma_{01}^* x_i + \gamma_{02}^* \varpi + \zeta_{0i}^*) \quad \text{(A3)}$$
$$= (\gamma_{01}^*)^2 \sigma_x^2 + (\gamma_{02}^*)^2 \sigma_\varpi^2 + \sigma_{\zeta_0^*}^2 + \gamma_{01}^* \gamma_{02}^* \mathrm{cov}(x_i, \varpi)$$
$$+ \gamma_{01}^* \mathrm{cov}(x_i, \zeta_{0i}^*) + \gamma_{02}^* \mathrm{cov}(\varpi, \zeta_{0i}^*)$$
$$= (\gamma_{01}^*)^2 \sigma_x^2 + (\gamma_{02}^*)^2 \sigma_\varpi^2 + \sigma_{\zeta_0^*}^2$$

The conditional variance of the latent intercept is:

$$\sigma_{\eta_0 | x}^2 = \mathrm{var}(\alpha_0^* + \gamma_{01}^* x_i + \gamma_{02}^* \varpi + \zeta_{0i}^* | x) = \mathrm{var}(\gamma_{02}^* \varpi + \zeta_{0i}^* | x) \quad \text{(A4)}$$
$$= (\gamma_{02}^*)^2 \sigma_\varpi^2 + \sigma_{\zeta_0^*}^2$$

Given (3), the unconditional variance of the latent slope is:

$$\sigma_{\eta_1}^2 = \mathrm{var}(\alpha_1^* + \gamma_{11}^* x_i + \gamma_{12}^* \varpi + \zeta_{1i}^*) \quad \text{(A5)}$$
$$= (\gamma_{11}^*)^2 \sigma_x^2 + \sigma_\varpi^2 (\gamma_{12}^*)^2 + \sigma_{\zeta_1^*}^2 + \gamma_{11}^* \gamma_{12}^* \mathrm{cov}(x_i, \varpi)$$
$$+ \gamma_{11}^* \mathrm{cov}(x_i, \zeta_{1i}^*) + \gamma_{12}^* \mathrm{cov}(\varpi, \zeta_{1i}^*)$$
$$= (\gamma_{11}^*)^2 \sigma_x^2 + (\gamma_{12}^*)^2 \sigma_\varpi^2 + \sigma_{\zeta_1^*}^2$$

The conditional variance of latent slope is:

$$\sigma_{\eta_1 | x}^2 = \mathrm{var}(\alpha_1^* + \gamma_{11}^* x_i + \gamma_{12}^* \varpi + \zeta_{1i}^* | x_i) \quad \text{(A6)}$$
$$= (\gamma_{12}^*)^2 \mathrm{var}(\varpi | x_i) + \mathrm{var}(\zeta_{1i}^* | x_i) = (\gamma_{12}^*)^2 \sigma_\varpi^2 + \sigma_{\zeta_1^*}^2$$

Next, we compute the unconditional covariance between the latent growth factors and ϖ.

$$\sigma_{\eta_0, \varpi} = \mathrm{cov}\big(\alpha_0^* + \gamma_{01}^* \, x_i + \gamma_{02}^* \, \varpi + \zeta_{0i}^* \quad, \varpi\big) \tag{A7}$$

$$= \gamma_{01}^* \mathrm{cov}(x_i, \varpi) + \gamma_{02}^* \, \sigma_\varpi^2 + \mathrm{cov}(\zeta_{0i}^*, \quad \varpi) = \gamma_{02}^* \, \sigma_\varpi^2$$

$$\sigma_{\eta_1, \varpi} = \mathrm{cov}\big(\alpha_1^* + \gamma_{11}^* \, x_i + \gamma_{12}^* \, \varpi + \zeta_{1i}^*, \varpi\big) \tag{A8}$$

$$= \gamma_{11}^* \mathrm{cov}(x_i, \varpi) + \gamma_{12}^* \, \sigma_\varpi^2 + \mathrm{cov}(\zeta_{1i}^*, \varpi) = \gamma_{12}^* \, \sigma_\varpi^2$$

Next, we compute the conditional and unconditional correlation between the latent growth factors and ϖ. Given (A3) and (A7), the unconditional correlation between the latent intercept and ϖ is:

$$\rho_{\eta_0, \varpi} = \frac{\sigma_{\eta_0, \varpi}}{\sigma_{\eta_0} \, \sigma_\varpi} = \frac{\gamma_{02}^* \, \sigma_\varpi}{\sqrt{(\gamma_{01}^*)^2 \, \sigma_x^2 + \sigma_\varpi^2 (\gamma_{02}^*)^2 + \sigma_{\zeta_0^*}^2}} = \frac{\gamma_{02}^* \, \sigma_\varpi}{\sigma_{\eta_0}} \tag{A9}$$

Given (A1) and (A4), the conditional correlation between the latent intercept and ϖ is:

$$\rho_{\eta_0, \varpi|x} = \frac{\sigma_{\eta_0, \varpi|x}}{\sigma_{\eta_0|x} \, \sigma_\varpi} = \frac{\gamma_{02}^* \, \sigma_\varpi}{\sqrt{\sigma_\varpi^2 \, (\gamma_{02}^*)^2 + \sigma_{\zeta_0^*}^2}} = \frac{\gamma_{02}^* \, \sigma_\varpi}{\sigma_{\eta_0|x}}$$

Given (A5) and (A8), the unconditional correlation between the latent slope and ϖ is:

$$\rho_{\eta_1, \varpi} = \frac{\sigma_{\eta_1, \varpi}}{\sigma_{\eta_1} \, \sigma_\varpi} = \frac{\gamma_{12}^* \, \sigma_\varpi}{\sqrt{(\gamma_{11}^*)^2 \, \sigma_x^2 + \sigma_\varpi^2 (\gamma_{12}^*)^2 + \sigma_{\zeta_1^*}^2}} = \frac{\gamma_{12}^* \, \sigma_\varpi}{\sigma_{\eta_1}} \tag{A10}$$

Given (A2) and (A6), the conditional correlation between the latent slope and ϖ is:

$$\rho_{\eta_1, \varpi|x} = \frac{\sigma_{\eta_1, \varpi|x}}{\sigma_{\eta_1|x} \, \sigma_\varpi} = \frac{\gamma_{12}^* \, \sigma_\varpi}{\sqrt{\sigma_\varpi^2 \, (\gamma_{12}^*)^2 + \sigma_{\zeta_1^*}^2}} = \frac{\gamma_{12}^* \, \sigma_\varpi}{\sigma_{\eta_1|x}}$$

Next, we derive $\sigma_{y, \varpi}$ and $\rho_{y, \varpi}$.

$$\sigma_{y,\varpi} = \text{cov}(\nu^* + \kappa^* \ x_i + \ \lambda_1^* \ \eta_{i0} + \lambda_2^* \ \eta_{i1} + \ \lambda_3^*\varpi + \ \varepsilon_i^*, \varpi) = \lambda_1^*\text{cov}(\ \eta_{i0}, \varpi) \quad \text{(A11)}$$

$$+ \ \lambda_2^*\text{cov}(\ \eta_{i1}, \varpi) + \ \lambda_3^* \ \sigma_\varpi^2 = (\ \lambda_1^* \ \gamma_{02}^* + \ \lambda_2^* \ \gamma_{12}^* + \ \lambda_3^*)\sigma_\varpi^2$$

$$\rho_{y,\varpi} = \frac{\sigma_{y_i,\varpi}}{\sigma_y \ \sigma_\varpi} \ = \frac{\sigma_\varpi}{\sigma_y}(\ \lambda_1^* \ \gamma_{02}^* + \ \lambda_2^* \ \gamma_{12}^* + \ \lambda_3^*)$$

Finally, isolating γ_{02}^* in (A9), we arrive at the expression in (9). Isolating γ_{12}^* in (A10), we arrive at the expression in (10). Isolating λ_3^* in (A11) would result in (11).

In addition, we compute the expected valued for the latent growth factors and y_i. We use the expected values to compare the respective residuals with the ones from the correlated augmented model in the next section.

$$E(\eta_{0i} \ | \ x_i) = E[E(\eta_{0i} \ | \ x_i, \varpi)] = E[E(\alpha_0^* + \gamma_{01}^* \ x_i + \gamma_{02}^* \ \varpi + \ \zeta_{0i}^* \ |x_i, \varpi)] = \alpha_0^* \quad \text{(A12)}$$
$$+ \gamma_{01}^* \ x_i + E[\gamma_{02}^* \ \varpi + E(\ \zeta_{0i}^* \ |x_i, \varpi)] = \alpha_0^* + \gamma_{01}^* \ x_i$$

$$E(\eta_{1i} \ | \ x_i) = E[E(\eta_{1i} \ | \ x_i, \varpi)] \quad \text{(A13)}$$
$$= E[E(\alpha_1^* + \gamma_{11}^* \ x_i + \gamma_{12}^* \ \varpi + \ \zeta_{1i}^* \ |x_i, \varpi)]$$
$$= \ \alpha_1^* + \gamma_{11}^* \ x_i + E[\gamma_{12}^* \ \varpi + E(\ \zeta_{1i}^* \ |x_i, \varpi)] = \ \alpha_1^* + \gamma_{11}^* \ x_i$$

$$E(y_i|x_i) = E(\nu^* + \kappa^* \ x_i + \ \lambda_1^* \ \eta_{i0} + \lambda_2^* \ \eta_{i1} + \ \lambda_3^* \ \varpi + \ \varepsilon_i^* \ | \ x_i, \ \eta_{0i}, \eta_{1i}) \quad \text{(A14)}$$
$$= \ \nu^* + \kappa^* \ x_i + \ \lambda_1^* \ \eta_{i0} + \lambda_2^* \ \eta_{i1}$$

Correlated Augmented Model

First, we derive the mean and variance of the mediators (growth factors) for the correlated augmented model and then compare them with the corresponding expected values from the latent augmented model. To show equivalency, the parameters from the correlated augmented model must to be a function of the corresponding parameters from the latent augmented model.

$$E(\eta_{0i} \ | \ x_i) = E[E(\eta_{0i} \ | \ x_i, \varpi)] = E[E(\alpha_0^{**} + \gamma_{01}^{**} \ x_i + \ \zeta_{0i}^{**} \ |x_i, \varpi)] = \alpha_0^{**} \quad \text{(A15)}$$
$$+ \gamma_{01}^{**} \ x_i + E[E(\ \zeta_{0i}^{**} \ |x_i, \varpi)] = \alpha_0^{**} + \gamma_{01}^{**} \ x_i + E(\zeta_{0i}^{**}|\varpi)$$
$$= \alpha_0^{**} + \gamma_{01}^{**} \ x_i$$

$$E(\eta_{1i} \mid x_i) = E[E(\eta_{1i} \mid x_i, \varpi)]$$
$$= E[E(\alpha_1^{**} + \gamma_{11}^{**} x_i + \zeta_{1i}^{**} \mid x_i, \varpi)] \quad \quad \text{(A16)}$$
$$= \alpha_1^{**} + \gamma_{11}^{**} x_i + E[E(\zeta_{1i}^{**} \mid x_i, \varpi)] = \alpha_1^{**} + \gamma_{11}^{**} x_i$$

$$E(y_i|x_i) = E(\nu^{**} + \kappa^{**} x_i + \lambda_1^{**} \eta_{0i} + \lambda_2^{**} \eta_{1i} + \varepsilon_y^{**} \mid x_i, \ \eta_{0i}, \eta_{1i}) \quad \text{(A17)}$$
$$= \nu^{**} + \kappa^{**} x_i + \lambda_1^{**} \eta_{0i} + \lambda_2^{**} \eta_{1i}$$

Comparing the expected values of the correlated augmented model with the corresponding expected values from the latent augmented model, one can see that there is a one-to-one relationship between the two model parameters. Further, we can use the expected values to isolate the residuals for latent growth factors and y_i in each model. Then, comparing the corresponding residuals from the two models, the following holds:

$$\zeta_{0i}^{**} = \gamma_{02}^* \varpi + \zeta_{0i}^*$$

$$\zeta_{1i}^{**} = \gamma_{12}^* \varpi + \zeta_{1i}^*$$

$$\varepsilon_i^{**} = \lambda_3^* \varpi + \varepsilon_i^*$$

We use the relationships between the residuals to compute the covariance and ultimately correlation between residuals in the correlated augmented shown (16)–(19) in terms of the parameters from the latent augmented model:

$$\sigma_{\varepsilon_y^{**}, \zeta_0^{**}} = \text{cov}\left(\varepsilon_i^{**}, \ \zeta_0^{**}\right) = \text{cov}(\varepsilon_i^{**}, \ \zeta_0^{**}| \ \eta_{i0}, \ \eta_{10}, x) = \text{cov}\left(\lambda_3^* \varpi + \ \varepsilon_i^*, \gamma_{02}^* \varpi + \ \zeta_{0i}^*\right)$$
$$= \gamma_{02}^* \ \lambda_3^* \ \sigma_\varpi^2$$

$$\sigma_{\varepsilon_y^{**}, \zeta_1^{**}} = \text{cov}\left(\varepsilon_i^{**}, \zeta_1^{**}\right) = \text{cov}(\varepsilon_i^{**}, \ \zeta_1^{**}| \ \eta_{i0}, \ \eta_{10}, x) = \text{cov}\left(\lambda_3^* \varpi + \ \varepsilon_i^*, \ \gamma_{12}^* \varpi + \ \zeta_{1i}^*\right)$$
$$= \gamma_{12}^* \ \lambda_3^* \ \sigma_\varpi^2$$

Next, we derive the following variances:

$$\sigma_{\varepsilon_y^{**}}^2 = \left(\lambda_3^*\right)^2 \ \sigma_\varpi^2 + \sigma_{\varepsilon_y^*}^2$$

$$\sigma^2_{\zeta_0^{**}} = \left(\gamma^*_{02}\right)^2 \; \sigma^2_\varpi + \sigma^2_{\zeta_0^*}$$

$$\sigma^2_{\zeta_1^{**}} = \; \left(\gamma^*_{12}\right)^2 \; \sigma^2_\varpi + \sigma^2_{\zeta_1^*}$$

Finally, the derive the following correlations shown in (18) and (19):

$$\rho_{\varepsilon^{**}_y, \zeta_0^{**}} = \frac{\sigma_{\varepsilon^{**}_y, \zeta_0^{**}}}{\sigma_{\varepsilon^{**}_y} \; \sigma_{\zeta_0^{**}}} = \frac{\gamma^*_{02} \lambda^*_3 \; \sigma^2_\varpi}{\sqrt{\left(\lambda^*_3\right)^2 \; \sigma^2_\varpi + \sigma^2_{\varepsilon^*_y}} \; \sqrt{\left(\gamma^*_{02}\right)^2 \; \sigma^2_\varpi + \sigma^2_{\zeta_0^*}}}$$

$$\rho_{\varepsilon^{**}_y, \zeta_1^{**}} = \frac{\sigma_{\varepsilon^{**}_y, \zeta_1^{**}}}{\sigma_{\varepsilon^{**}_y} \; \sigma_{\zeta_1^{**}}} = \frac{\gamma^*_{12} \lambda^*_3 \; \sigma^2_\varpi}{\sqrt{\left(\lambda^*_3\right)^2 \; \sigma^2_\varpi + \sigma^2_{\varepsilon^*_y}} \; \sqrt{\left(\gamma^*_{12}\right)^2 \; \sigma^2_\varpi + \sigma^2_{\zeta_1^*}}}$$

Below we derive the relationships between the covariance between residuals of the growth factors in the correlated augmented model and the corresponding covariance in the latent augmented model.

$$\sigma_{\zeta_1^{**}, \zeta_0^{**}} = \text{cov}(\zeta_1^{**}, \; \zeta_0^{**} \; | \; \eta_{i0}, \quad \eta_{10}, \boldsymbol{x}) = \text{cov}\left(\gamma^*_{12}\varpi + \; \zeta^*_{1i}, \; \gamma^*_{02}\varpi + \zeta^*_{0i}\right)$$
$$= \; \gamma^*_{02} \; \gamma^*_{12} \; \sigma^2_\varpi + \sigma_{\zeta_0^*, \zeta_1^*}$$

This the expression shown in Equation (5).

References

Allison PD (2005). Fixed effects regression methods for longitudinal data using SAS (1st ed.). Cary, NC: SAS Publishing.

Alvarez AN, & Juang LP (2010). Filipino Americans and racism: A multiple mediation model of coping. Journal of Counseling Psychology, 57, 167–178. 10.1037/a0019091 [PubMed: 21133568]

Baraldi AN, & Enders CK (2010). An introduction to modern missing data analyses. Journal of School Psychology, 48, 5–37. 10.1016/j.jsp.2009.10.001 [PubMed: 20006986]

Bartholomew DJ, Knott M, & Moustaki I (2011). Latent variable models and factor analysis: a unified approach (3rd ed). Chichester, West Sussex, UK: Wiley.

Bind M-AC, VanderWeele TJ, Coull BA, & Schwartz JD (2016). Causal mediation analysis for longitudinal data with exogenous exposure. Biostatistics, 17, 122–134. 10.1093/biostatistics/kxv029 [PubMed: 26272993]

Chen F, Bollen KA, Paxton P, Curran PJ, & Kirby JB (2001). Improper solutions in structural equation models: Causes, consequences, and strategies. Sociological Methods & Research, 29(4), 468–508. 10.1177/0049124101029004003

Chen Z, & Wang H (2017). Abusive supervision and employees' job performance: A multiple mediation model. Social Behavior and Personality, 45(5), 845–858. 10.2224/sbp.5657

Cohen J (1988). Statistical power analysis for the behavioral sciences (2nd ed.). Hillsdale, N.J.: Erlbaum.

Cole DA, Ciesla JA, & Steiger JH (2007). The insidious effects of failing to include design-driven correlated residuals in latent-variable covariance structure analysis. Psychological Methods, 12, 381–398. 10.1037/1082-989X.12.4.381 [PubMed: 18179350]

Cox MG, Kisbu-Sakarya Y, Mio evi M, & MacKinnon DP (2013). Sensitivity plots for confounder bias in the single mediator model. Evaluation Review, 37, 405–431. 10.1177/0193841X14524576 [PubMed: 24681690]

Curran PJ (2003). Have multilevel models been structural equation models all along? Multivariate Behavioral Research, 38, 529–569. 10.1207/s15327906mbr3804_5 [PubMed: 26777445]

Daniel RM, De Stavola BL, Cousens SN, & Vansteelandt S (2015). Causal mediation analysis with multiple mediators. Biometrics, 71(1), 1–14. 10.1111/biom.12248 [PubMed: 25351114]

Demiray B, & Janssen SMJ (2015). The self-enhancement function of autobiographical memory. Applied Cognitive Psychology, 29, 49–60. 10.1002/acp.3074

Dillon WR, Kumar A, & Mulani N (1987). Offending estimates in covariance structure analysis: Comments on the causes of and solutions to Heywood cases. Psychological Bulletin, 101, 126–135. 10.1037/0033-2909.101.1.126

Flannery BA, Volpicelli JR, & Pettinati HM (1999). Psychometric properties of the Penn Alcohol Craving Scale. Alcoholism: Clinical and Experimental Research, 23(8), 1289–1295.

Fritz MS, Kenny DA, & MacKinnon DP (2016). The combined effects of measurement error and omitting confounders in the single-mediator model. Multivariate Behavioral Research, 51(5), 681–697. 10.1080/00273171.2016.1224154 [PubMed: 27739903]

Fritz MS, & MacKinnon DP (2007). Required sample size to detect the mediated effect. Psychological Science, 18, 233–239. 10.1111/j.1467-9280.2007.01882.x [PubMed: 17444920]

Grimm KJ, Ram N, & Estabrook R (2017). Growth modeling: Structural equation and multilevel modeling approaches New York, NY: Guilford Press.

Hallgren KA, Wilson AD, & Witkiewitz K (2018). Advancing analytic approaches to address key questions in mechanisms of behavior change research. Journal of Studies on Alcohol and Drugs, 79, 182–189. 10.15288/jsad.2018.79.182 [PubMed: 29553344]

Harring JR, McNeish DM, & Hancock GR (2017). Using phantom variables in structural equation modeling to assess model sensitivity to external misspecification. Psychological Methods, 22, 616–631. 10.1037/met0000103 [PubMed: 29265846]

Holland PW (1988). Causal Inference, path analysis, and recursive structural equations models. Sociological Methodology, 18, 449–484. 10.2307/271055

Hoyle RH (Ed.). (2012). Handbook of structural equation modeling New York, NY: Guilford Press.

Imai K, Keele L, & Tingley D (2010). A general approach to causal mediation analysis. Psychological Methods, 15, 309–334. 10.1037/a0020761 [PubMed: 20954780]

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

Imai K, & Yamamoto T (2013). Identification and sensitivity analysis for multiple causal mechanisms: revisiting evidence from framing experiments. Political Analysis, 21, 141–171.

James LR, & Brett JM (1984). Mediators, moderators, and tests for mediation. Journal of Applied Psychology, 69, 307–321. 10.1037/0021-9010.69.2.307

Judd CM, & Kenny DA (1981). Process analysis. Evaluation Review, 5, 602–619. 10.1177/0193841X8100500502

Kenny DA (2004). Correlation and causality (Revised) New York: Wiley Retrieved from http://davidakenny.net/doc/cc_v1.pdf

Kenny DA, & Milan S (2012). Identification: A non-technical discussion of a technical issue. In Hoyle RH (Ed.), Handbook of structural equation modeling (pp. 145–163). New York, NY: Guilford Press.

Kershaw KN, Mezuk B, Abdou CM, Rafferty JA, & Jackson JS (2010). Socioeconomic position, health behaviors, and C-reactive protein: A moderated-mediation analysis. Health Psychology, 29, 307–316. 10.1037/a0019286 [PubMed: 20496985]

Kline RB (2016). Principles and practice of structural equation modeling (Fourth edition). New York: Guilford.

Kolenikov S, & Bollen KA (2012). Testing negative error variances: Is a Heywood case a symptom of misspecification? Sociological Methods & Research, 41(1), 124–167. 10.1177/0049124112442138

Kranzler HR, Armeli S, Covault J, & Tennen H (2013). Variation in OPRM1 moderates the effect of desire to drink on subsequent drinking and its attenuation by naltrexone treatment: Naltrexone pharmacogenetics. Addiction Biology, 18, 193–201. 10.1111/j.1369-1600.2012.00471.x [PubMed: 22784013]

Kreft IGG, de Leeuw J, & Aiken LS (1995). The effect of different forms of centering in hierarchical linear models. Multivariate Behavioral Research, 30, 1–21. 10.1207/s15327906mbr3001_1 [PubMed: 26828342]

Leung C-K, & Lam K (1975). A note on the geometric representation of the correlation coefficients. The American Statistician, 29, 128–130. 10.2307/2683440

Ma Y, Cheng W, Ribbens BA, & Zhou J (2013). Linking ethical leadership to employee creativity: Knowledge sharing and self efficacy as mediators. Social Behavior and Personality, 41, 1409–1419. 10.2224/sbp.2013.41.9.1409

MacKinnon DP (2008). Introduction to statistical mediation analysis New York, NY: Erlbaum.

MacKinnon DP, Fairchild AJ, & Fritz MS (2007). Mediation analysis. Annual Review of Psychology, 58, 593–614. 10.1146/annurev.psych.58.110405.085542

MacKinnon DP, & Pirlott AG (2015). Statistical approaches for enhancing causal interpretation of the M to Y relation in mediation analysis. Personality and Social Psychology Review, 19, 30–43. 10.1177/1088868314542878 [PubMed: 25063043]

Mann K, Roos CR, Hoffmann S, Nakovics H, Leménager T, Heinz A, & Witkiewitz K (2018). Precision medicine in alcohol dependence: A controlled trial testing pharmacotherapy response among reward and relief drinking phenotypes. Neuropsychopharmacology, 43, 891–899. 10.1038/npp.2017.282 [PubMed: 29154368]

McNeish D, & Kelley K (2018). Fixed effects models versus mixed effects models for clustered data: Reviewing the approaches, disentangling the differences, and making recommendations. Psychological Methods 10.1037/met0000182

Muthén B, & Asparouhov T (2015). Causal effects in mediation modeling: An introduction with applications to latent variables. Structural Equation Modeling: A Multidisciplinary Journal, 22, 12–23. 10.1080/10705511.2014.935843

Muthén LK, & Muthén BO (2017). Mplus User's Guide (Version 8) Los Angeles, CA: Muthén & Muthén.

Pearl J (2011). The mediation formula: A guide to the assessment of causal pathways in nonlinear models. In Berzuini C, Dawid P, & Bernardinell L (Eds.), Causality: Statistical Perspectives and Applications (pp. 151–179.). Hoboken, N.J.

Pearl J (2014). Interpretation and identification of causal mediation. Psychological Methods, 19, 459–481. 10.1037/a0036434 [PubMed: 24885338]

Preacher KJ (2015). Advances in mediation analysis: A survey and synthesis of new developments. Annual Review of Psychology, 66, 825–852. 10.1146/annurev-psych-010814-015258

Priesemuth M, Schminke M, Ambrose ML, & Folger R (2014). Abusive supervision climate: A multiple-mediation model of its impact on group outcomes. Academy of Management Journal, 57, 1513–1534.

R Development Core Team. (2017). R: A Language and Environment for Statistical Computing (Version 3.4.3) Vienna, Austria: R Foundation for Statistical Computing Retrieved from http://www.R-project.org/

Raudenbush SW, & Bryk AS (2002). Hierarchical linear models: Applications and data analysis methods (2nd ed.). Thousand Oaks, CA: Sage.

Rindskopf D (1983). Parameterizing inequality constraints on unique variances in linear structural models. Psychometrika, 48(1), 73–83. 10.1007/BF02314677

Rindskopf D (1984). Structural equation models: Empirical identification, Heywood cases, and related problems. Sociological Methods & Research, 13, 109–119. 10.1177/0049124184013001004

Robins JM, & Greenland S (1992). Identifiability and exchangeability for direct and indirect effects. Epidemiology, 3(2), 143–155. 10.1097/00001648-199203000-00013 [PubMed: 1576220]

Rousseeuw PJ, & Molenberghs G (1994). The shape of correlation matrices. The American Statistician, 48, 276–279. 10.2307/2684832

Rubin DB (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. Journal of Educational Psychology, 66(5), 688–701. 10.1037/h0037350

Rubin DB (1978). Bayesian inference for causal effects: The role of randomization. The Annals of Statistics, 6, 34–58.

Schacht JP, Randall PK, Latham PK, Voronin KE, Book SW, Myrick H, & Anton RF (2017). Predictors of naltrexone response in a randomized trial: Reward-related rrain activation, OPRM1 genotype, and smoking status. Neuropsychopharmacology, 42(13), 2640–2653. 10.1038/npp. 2017.74 [PubMed: 28409564]

Selig JP, & Preacher KJ (2009). Mediation models for longitudinal data in developmental research. Research in Human Development, 6, 144–164. 10.1080/15427600902911247

Singer JD, & Willett JB (2003). Applied longitudinal data analysis modeling change and event occurrence Oxford, MA: Oxford University Press.

Snijders TAB, & Bosker RJ (2011). Multilevel analysis: An introduction to basic and advanced multilevel modeling (2nd ed.). Thousand Oaks, CA: SAGE.

Strelan P, Karremans JC, & Krieg J (2017). What determines forgiveness in close relationships? The role of post-transgression trust. British Journal of Social Psychology, 56, 161–180. 10.1111/bjso. 12173 [PubMed: 27862012]

Talloen W, Moerkerke B, Loeys T, De Naeghel J, Van Keer H, & Vansteelandt S (2016). Estimation of indirect effects in the presence of unmeasured confounding for the mediator–outcome relationship in a multilevel 2–1–1 mediation model. Journal of Educational and Behavioral Statistics, 41, 359–391. 10.3102/1076998616636855

Tofighi D, & Kelley K (2016). Assessing omitted confounder bias in multilevel mediation models. Multivariate Behavioral Research, 51, 86–105. 10.1080/00273171.2015.1105736 [PubMed: 26881959]

Tofighi D, & MacKinnon DP (2011). RMediation: An R package for mediation analysis confidence intervals. Behavior Research Methods, 43, 692–700. 10.3758/s13428-011-0076-x [PubMed: 21487904]

Tofighi D, West SG, & MacKinnon DP (2013). Multilevel mediation analysis: The effects of omitted variables in the 1–1–1 model. British Journal of Mathematical and Statistical Psychology, 66, 290–307. 10.1111/j.2044-8317.2012.02051.x [PubMed: 22594884]

Valente MJ, Pelham WEI, Smyth H, & MacKinnon DP (2017). Confounding in statistical mediation analysis: What it is and how to address it. Journal of Counseling Psychology, 64(6), 659–671. 10.1037/cou0000242 [PubMed: 29154577]

VanderWeele TJ (2010). Bias formulas for sensitivity analysis for direct and indirect effects. Epidemiology, 21, 540–551. 10.1097/EDE.0b013e3181df191c [PubMed: 20479643]

VanderWeele TJ (2015). Explanation in causal inference: Methods for mediation and interaction New York, NY: Oxford University Press.

Villegas-Gold R, & Yoo HC (2014). Coping with discrimination among Mexican American college students. Journal of Counseling Psychology, 61(3), 404–413. 10.1037/a0036591 [PubMed: 25019544]

von Soest T, & Hagtvet KA (2011). Mediation analysis in a latent growth curve modeling framework. Structural Equation Modeling: A Multidisciplinary Journal, 18, 289–314. 10.1080/10705511.2011.557344

West SG, Finch JF, & Curran PJ (1995). Structural equation models with nonnormal variables: Problems and remedies. In Hoyle RH (Ed.), Structural equation modeling: Concepts, issues, and applications (pp. 56–75). Thousand Oaks, CA, US: Sage.

Willett JB, & Bub KL (2005). Structural Equation Modeling: Latent Growth Curve Analysis. In Everitt BS & Howell DC (Eds.), Encyclopedia of Statistics in Behavioral Science Chichester, UK: John Wiley & Sons, Ltd 10.1002/0470013192.bsa599
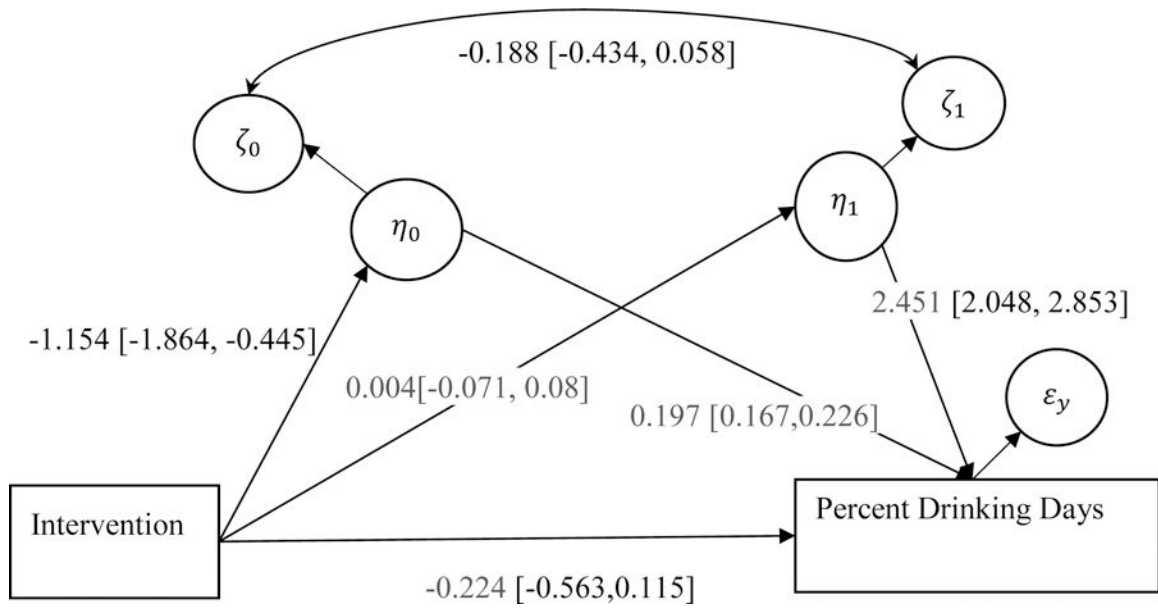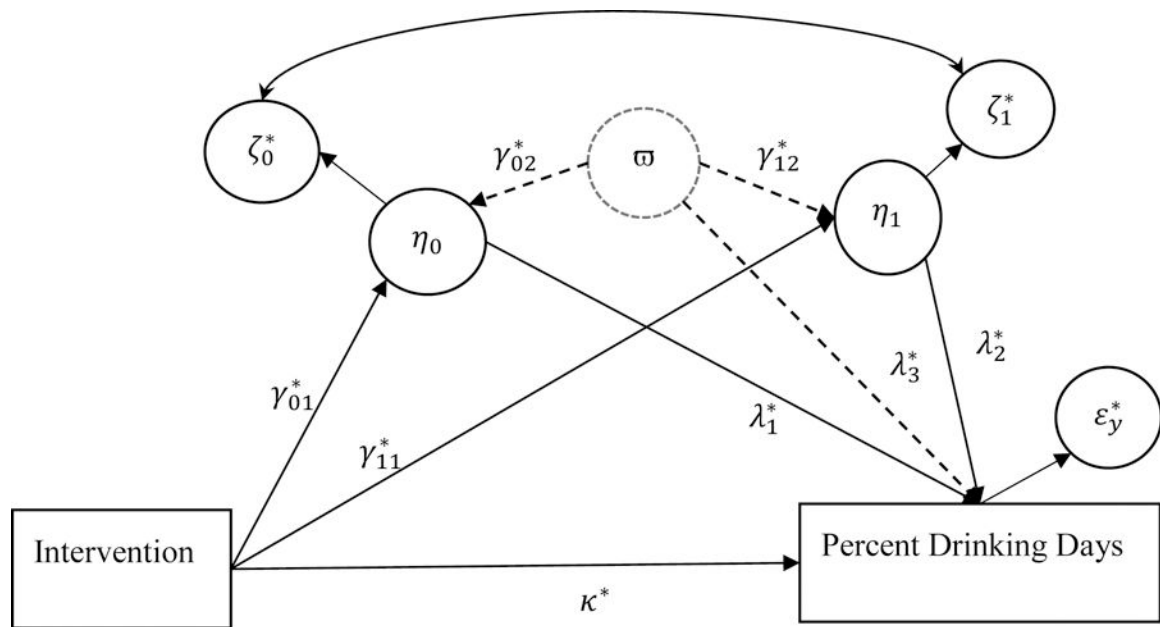
Figure 1.
Latent growth mediation model. Indirect effect of intervention (naltrexone vs non-naltrexone) treatment on percent drinking days at week 16 is assumed to be mediated via latent intercept (mean alcohol craving in week 4) and weekly growth rate of alcohol craving. The percent drinking days is divided by 10 to make its dispersion smaller and more comparable to rest of the data. For the indirect effect involving latent intercept, B (SE) = −0.227 (0.073), 97.5% CI [−0.398, −0.067]; for the indirect effect involving latent slope, B (SE) = 0.01 (0.094), 97.5% CI [−0.206, 0.227]. A solid straight arrow shows the effect of a variable at the origin on the variable at the end of the arrow. A solid double-headed curved arrow shows covariance between two variables. Observed variables associated with the latent intercept and slopes are excluded for simplicity.
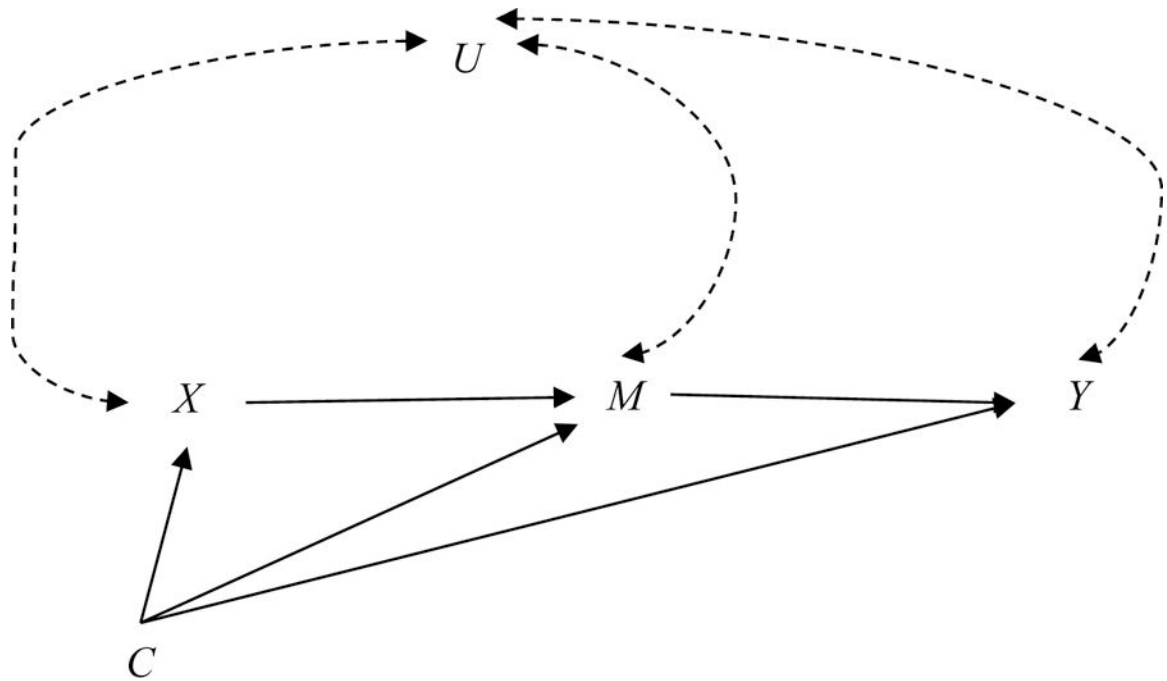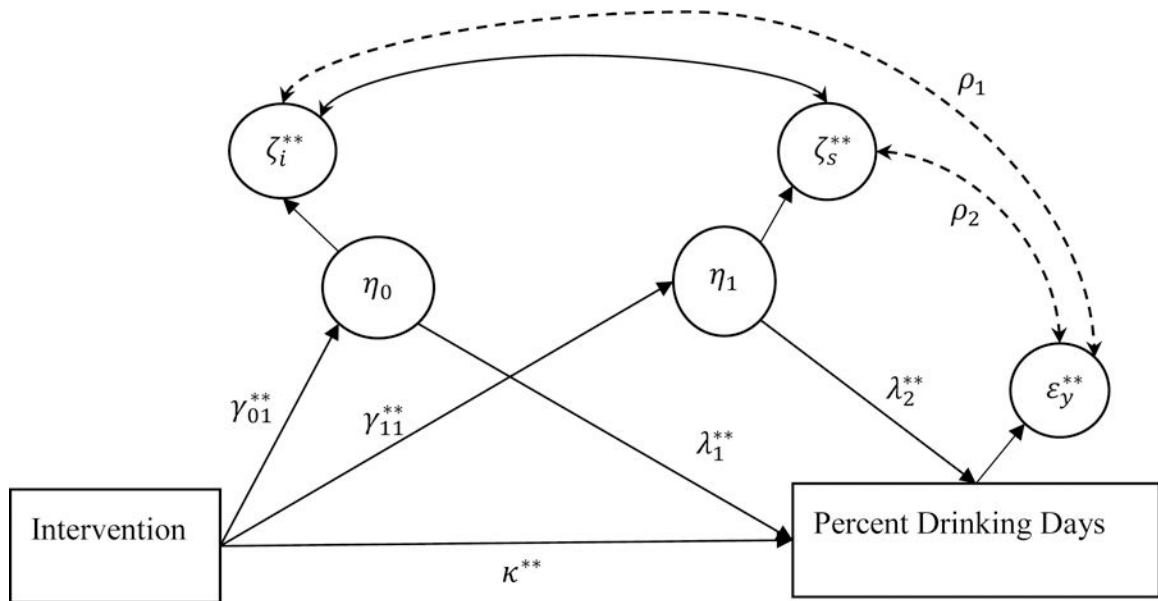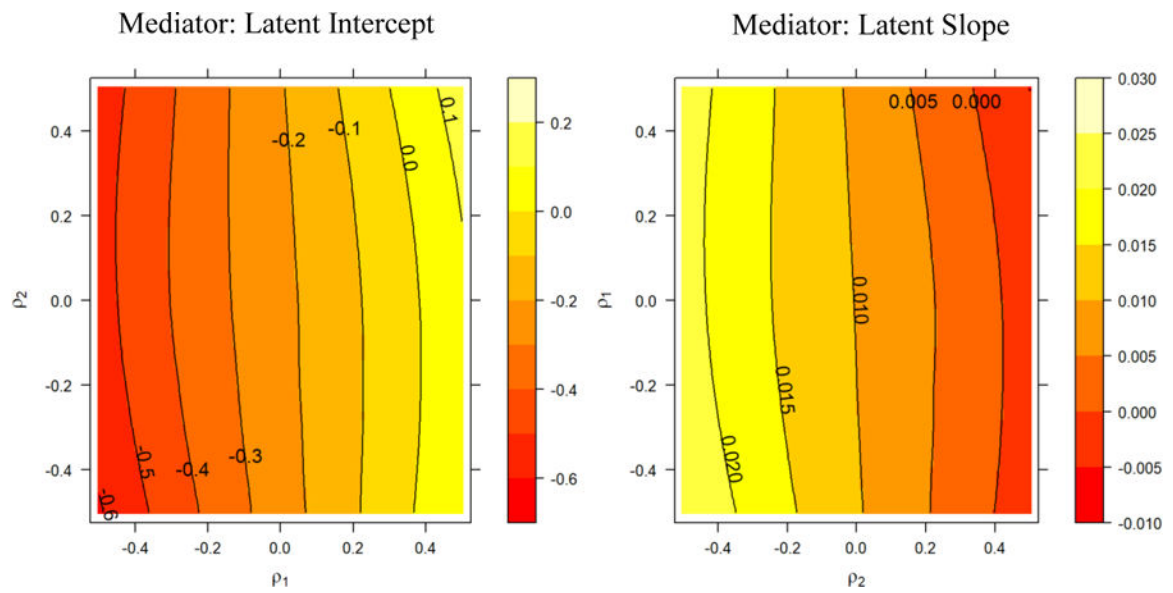
Figure 2.
A general single mediator model. All the variables can be observed or latent. X, M, and Y denotes antecedent, mediator, and outcome variable, respectively. U denotes one or more omitted confounders. C denotes one or covariates (e.g., background variables) included in the model. We assume that C is measured before X, which is measured before M, and which is measured before Y. A solid arrow shows the effect of a variable at the origin on the variable at the end of the arrow. A dashed double headed curved arrow illustrates a confounder correlation (covariance) between U and another variable in the model.

Figure 3.
Latent augmented model. Indirect effect of intervention (naltrexone vs placebo) treatment on percent drinking days at week 16 is assumed to be mediated via latent intercept (mean alcohol craving at week 4) and weekly growth rate of alcohol craving from weeks 6 through 12.
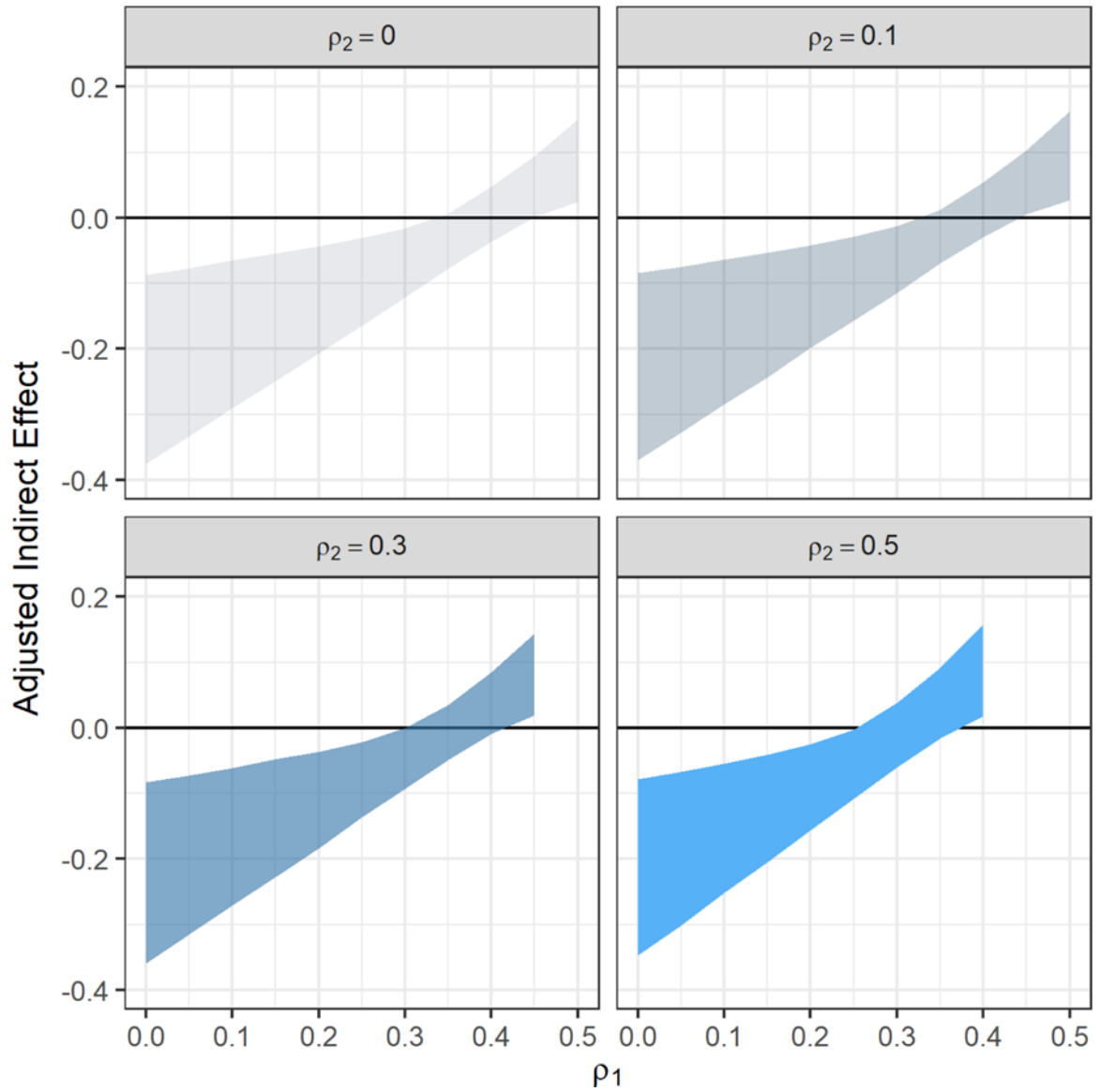
Figure 4.
Correlated augmented model. Indirect effect of intervention (naltrexone vs placebo)
treatment on percent drinking days at week 16 is assumed to be mediated via latent intercept
(mean alcohol craving at week 4) and weekly growth rate of alcohol craving from weeks 6
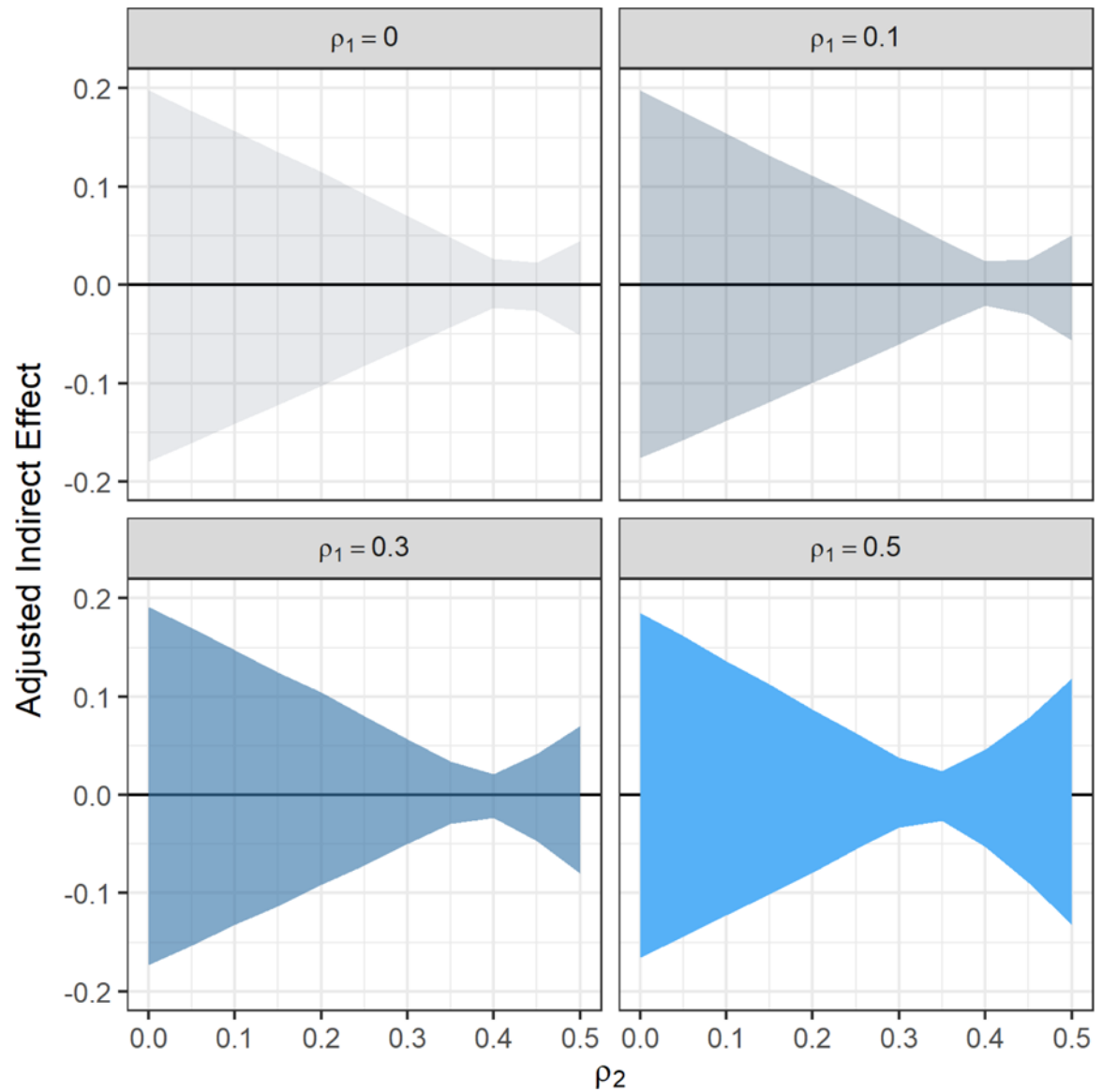through 12.

Figure 5.
Sensitivity contour plots for the indirect effects. Numbers adjacent to each contour line indicate indirect effect estimates. For the graph on the right, the independent variable, the mediator, and outcome variables are intervention, latent intercept, and heavy drinking as shown in Figure 4. For the graph on the left, the independent variable, the mediator, and outcome variables are intervention, latent slope, and drinking outcome. The confounder parameter ρ_1 is the correlation between the residuals associated with the latent intercept and outcome variable and the confounder parameter ρ_2 is the correlation between the residuals associated with latent slope and the outcome variable.

Figure 6.
Sensitivity confidence band graphs of 97.5% CIs for indirect effect through the mediator latent intercept. The x-axis on each panel graph shows the confounder parameter $\rho 1$, which is the correlation between the residuals associated with the latent intercept and outcome variable for the correlated augmented model in Figure 4. The confounder parameter, $\rho 2$, on top of each panel graph is the correlation between the residuals associated with latent slope and the outcome variable in Figure 4. The y-axis shows indirect effect estimate as a function of the confounder correlation parameters. The ribbon in each graph shows the 97.5% CI bands.

Figure 7.
Sensitivity confidence band graphs of 97.5% CIs for the indirect effect through the mediator latent slope. The *x*-axis on each panel graph shows the confounder parameter ρ2, which is the correlation between the residuals associated with the latent slope and outcome variable for the correlated augmented model in Figure 4. The confounder parameter, ρ1, on top of each panel graph is the correlation between the residuals associated with latent intercept and the outcome variable in Figure 4. The *y*-axis shows indirect effect estimate as a function of the confounder correlation parameters. The ribbon in each graph shows the 97.5% CI bands.

Table 1.

Summary of Mediation Models and Sensitivity Analysis

Study	Sensitivity analysis	Effect sizes for sensitivity perameter	Type of Mediation Model Used in Analysis			Statistical moduling
			One mediator	Two parallel mediators	Sequential mediators	
			Single level (independent) data			
Cox et al.(2013)	Yes	Yes	X→M→Y			SEM
Daniel et al. (2015)	Yes				X→M_1→M_2→M_3→Y	
Harring et al. (2017)	Yes	No	X→M→Y		X→M_1→M_2→Y	SEM
Imai, Keele, & Tingley (2010)	Yes	Yes	X→M→Y			GLM
Imai and Yamamato (2013)	Yes	Yes		X→($M_1 \perp M_2$)→Y	X→M_1→M_2→Y	GLM
Velante et al. (2017)	Yes	Yes	X→M→Y			SEM
			Multilevel (correlated) data			
Bind et al. (2016)	No		Within X→M→Y			Generalized mixed model
Talloen et al. (2016)	Yes[a]	Yes	Within X→M→Y			Fixed effects and Linear mixed model
Tofighi and Kelly (2016)	Yes	Yes	Both Between and Within X→M→Y			SEM
Current study	Yes	Yes		BetweenX→($M_1 \leftrightarrow M_2$)→Y		SEM

Note. GLM= generalized linear model; SEM=structural equation model. *X*, *M*, and *Y* denote independent, mediator, and outcome variables. The parentheses around the mediators identify parallel mediators. Symbol $\perp$ indicates that the parallel mediators are independent and symbol $\leftrightarrow$ indicates the parallel mediators are correlated.

[a]Used sensitivity analysis by Imai et al. (2010).