CHAPTER 22

Mediation/Indirect Effects in Structural Equation Modeling

Oscar Gonzalez
Matthew J. Valente
Jeewon Cheong
David P. MacKinnon

Thus, the correlation between two variables is equal to the sum of the products of chains of path coefficients along all of the paths by which they are connected.

—Sewall Wright, 1921, p. 568

The development of mediation models goes hand in hand with the history of SEM (see Matsueda, Chapter 2, this volume), and its applications in the social sciences (Duncan, 1966; Kendall & Lazarsfeld, 1950; Lazarsfeld, 1955). Modern approaches to quantifying indirect effects or mediation began with Sewall Wright's work on path analysis (Wolfle, 1999; Wright, 1920, 1921). Investigating the relative influences of heredity and environment on the breeding of guinea pigs, he used a system of equations and a path diagram to display the causal relations among the variables and showed that the contribution of the intermediate variables connecting the two variables in a complex system of causally related variables was the product of the path coefficients in a chain of paths connecting the two variables (Wright, 1923). He further argued that path analysis was not a method to infer causation but a method to quantify the supposed causal relations. Causal sequences of the variables were already assumed, based on available information including theory, prior correlations, and results of prior experiments; path analysis was used to gauge the strength of the causal relations. The specification of causal relations among variables is still one of the main issues in the use of SEM and we discuss this issue in relation to indirect or mediation effects later in the chapter (for a more general treatment of causal inference in SEM and structural causal models, see Pearl, Chapter 3, this volume).

With the development of covariance structure modeling during the 1970s (Jöreskog, 1970, 1973; Wiley, 1973) and 1980s (Bentler, 1980), the multiple equation tradition in path analysis was combined with the psychometric tradition in its focus on measurement; thus, measurement and structural models were incorporated in the quantification and modeling of indirect effects. General methods to decompose the association between the antecedent and the dependent variables into direct and indirect effects were developed (Alwin & Hauser, 1975; Graff & Schmidt, 1982) for covariance structure models, and Sobel (1982) derived the standard error of these direct and indirect effects. These standard errors based on multivariate delta method (Sobel, 1982) have

been implemented in most of the SEM programs to compute confidence intervals and conduct significance tests for indirect effects. Methods to accommodate the non-normal distribution of indirect effects were also introduced (MacKinnon, Lockwood, & Williams, 2004) and are now part of several SEM programs.

DEFINITIONS

A "mediator," or "mediating variable," is defined as a third variable that intervenes in the causal pathway between an independent variable and a dependent variable, transmitting the effect of the independent variable on the dependent variable (Baron & Kenny, 1986; Last, 1988; Sobel, 1982). Statistically, a third variable is considered a significant mediator when the relation between the independent and the dependent variables is completely or partially accounted for by the third variable intermediate in the causal chain. Mediators are different from two other fundamental third-variable causal relations: confounders and colliders (Pearl, 2012). A confounder may explain the effect of the independent variable and the dependent variable because it is related to both variables. However, a mediator explicates the effect as an intermediate variable in the causal sequence, whereas a confounder causes both the independent and the dependent variables. A collider variable is caused by the independent and dependent variables and should not be conditioned on in the statistical analysis because adjusting for a collider will distort the relation between the independent and dependent variable. A collider differs from a mediator because it is not intermediate in the causal sequence. In many situations, it may be difficult to determine whether the third variable is a mediator, a confounder, or a collider (MacKinnon & Lamp, 2021); researchers will need to rely on theory for making such decisions. In this chapter, we assume that the third variable is a mediator, and the model specifying the causal sequence of the independent variables, the mediators, and the dependent variables is correct. However, most comprehensive models include examples of mediator, confounder, and collider third-variable effects. Another differentiation researchers often make is the naming of the variables involved in the causal sequence in a mediation model. Kenny, Kashy, and Bolger (1998) used initial variable, mediator, and outcome. James and Brett (1984) incorporated the time sequence of the variables by referring to them as the antecedent, mediating, and consequent variables. Although certain names for the variables in the mediation model fit specific modeling situations better, the names we use in this chapter are *independent variable, mediator*, and *dependent variable*.

MEDIATION MODELS

Our focus is on indirect or mediation effects estimated in the SEM framework. These models typically have more than one independent variable, mediating variable, or dependent variable; all or some of the variables are modeled as latent variables. Figure 22.1 presents a single mediator model with observed variables. The coefficients relating the exogenous variable to the endogenous variables are indicated as γ's, and the coefficients relating the endogenous variables to other endogenous variables are indicated as β's. Figure 22.1 also contains actual path coefficients between the variables, obtained by analyzing a real data set from 1,208 high school football players (Goldberg et al., 2000). The football players were randomly assigned to either treatment or control groups. The treatment program (ξ) was expected to increase peer influence on healthy diet (η_1), which then was expected to improve the athletes' nutrition behaviors (η_2). The relations between the independent variable, the mediator, and the dependent variables are specified in the structural relations as follows:

$$\eta = \mathbf{B}\eta + \mathbf{\Gamma}\xi + \zeta \qquad (22.1)$$

where η is a 2×1 vector representing the mediator and the dependent variable (i.e., endogenous variables) in Figure 22.1, ζ is a 2×1 vector of their residuals, and ξ is a scalar representing an independent variable (i.e., an exogenous variable). The $\mathbf{\Gamma}$ matrix is a 2×1 vector

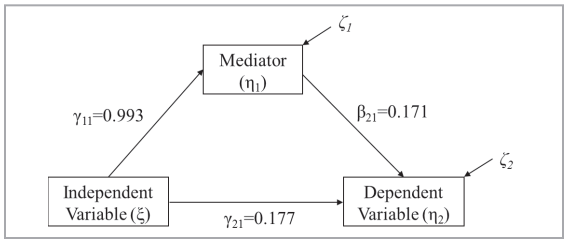

FIGURE 22.1. A single-mediator model with three manifest variables.

representing the relations of the independent variable to the mediator and the dependent variable, and the **B** matrix is a 2 × 2 matrix representing the relation between the mediator and the dependent variable. Specifically, the relations between the variables in Equation 22.1 are

$$\begin{bmatrix}\eta_1\\ \eta_2\end{bmatrix}=\begin{bmatrix}0 & 0\\ \beta_{21} & 0\end{bmatrix}\begin{bmatrix}\eta_1\\ \eta_2\end{bmatrix}+\begin{bmatrix}\gamma_{11}\\ \gamma_{21}\end{bmatrix}[\xi]+\begin{bmatrix}\varepsilon_1\\ \varepsilon_2\end{bmatrix}$$
$$=\begin{bmatrix}0 & 0\\ 0.171 & 0\end{bmatrix}\begin{bmatrix}\eta_1\\ \eta_2\end{bmatrix}+\begin{bmatrix}0.993\\ 0.177\end{bmatrix}[\xi]+\begin{bmatrix}\varepsilon_1\\ \varepsilon_2\end{bmatrix} \quad (22.2)$$

In models with manifest variables, it is assumed that each variable has been assessed appropriately. In other words, we assume that the scores on the variables are reliable (i.e., free of measurement error) and that they can be ascribed a valid interpretation (Gonzalez & MacKinnon, 2021). In the presence of measurement error, the relations between the variables can be attenuated as a function of the reliability of the predictor variable; consequently, the reliability of the mediator variable can lead to reduced mediation effects (Hoyle & Kenny, 1999). Also, if the representation of the mediating construct on the model is not appropriate (e.g., the mediator is only a facet of a broad construct, but the broad construct is in the mediation model), the mediation effects could go undetected (Gonzalez & MacKinnon, 2018). Introducing latent variables can improve estimation and interpretation of mediation effects by specifying a measurement model for each construct and separating measurement error from the true variance. Including measurement models for the constructs requires additional matrices specifying the relations between each indicator, and the relevant latent construct and matrices representing the remaining unexplained variability of each indicator.

Figure 22.2 depicts a two-mediator model with all latent variables, each with three indicators. For illustration, we used the data from 818 high school football players who participated in the Adolescents Training and Learning to Avoid Steroids (ATLAS) prevention program (Goldberg et al., 2000). It was hypothesized that athletic competence (ξ_1) would affect body image (η_1) and self-esteem (η_2), which then would affect the athlete's attitude toward anabolic steroids (η_3). In addition, body image (η_1) was hypothesized to affect self-esteem (η_2). All the parameter estimates are presented in Greek letters in Figure 22.2, with actual estimates of

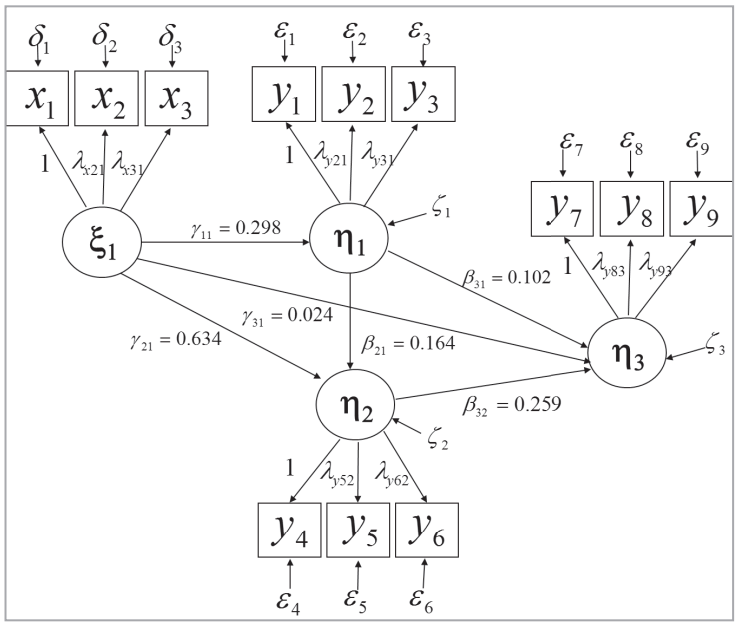

FIGURE 22.2. A two-mediator model with latent variables.

the relations between the latent variables. The measurement models in Figure 22.2 are specified as Equations 22.3 and 22.4, and the structural model is specified as Equations 22.5 and 22.6:

$$x = \Lambda_x \xi + \delta$$

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 1 \\ \lambda_{x_2 1} \\ \lambda_{x_3 1} \end{bmatrix} \begin{bmatrix} \xi_1 \end{bmatrix} + \begin{bmatrix} \delta_1 \\ \delta_2 \\ \delta_3 \end{bmatrix} \tag{22.3}$$

$$y = \Lambda_y \eta + \varepsilon$$

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \\ y_7 \\ y_8 \\ y_9 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ \lambda_{y_2 1} & 0 & 0 \\ \lambda_{y_3 1} & 0 & 0 \\ 0 & 1 & 0 \\ 0 & \lambda_{y52} & 0 \\ 0 & \lambda_{y62} & 0 \\ 0 & 0 & 1 \\ 0 & 0 & \lambda_{y_8 3} \\ 0 & 0 & \lambda_{y_9 3} \end{bmatrix} \begin{bmatrix} \eta_1 \\ \eta_2 \\ \eta_3 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \\ \varepsilon_5 \\ \varepsilon_6 \\ \varepsilon_7 \\ \varepsilon_8 \\ \varepsilon_9 \end{bmatrix} \tag{22.4}$$

$$\eta = \mathbf{B}\eta + \mathbf{\Gamma}\xi + \zeta \tag{22.5}$$

$$\begin{aligned} \begin{bmatrix} \eta_1 \\ \eta_2 \\ \eta_3 \end{bmatrix} &= \begin{bmatrix} 0 & 0 & 0 \\ \beta_{21} & 0 & 0 \\ \beta_{31} & \beta_{32} & 0 \end{bmatrix} \begin{bmatrix} \eta_1 \\ \eta_2 \\ \eta_3 \end{bmatrix} + \begin{bmatrix} \gamma_{11} \\ \gamma_{21} \\ \gamma_{31} \end{bmatrix} \begin{bmatrix} \xi_1 \end{bmatrix} + \begin{bmatrix} \zeta_1 \\ \zeta_2 \\ \zeta_3 \end{bmatrix} \\ &= \begin{bmatrix} 0 & 0 & 0 \\ 0.164 & 0 & 0 \\ 0.102 & 0.259 & 0 \end{bmatrix} \begin{bmatrix} \eta_1 \\ \eta_2 \\ \eta_3 \end{bmatrix} + \begin{bmatrix} 0.298 \\ 0.634 \\ 0.024 \end{bmatrix} \begin{bmatrix} \xi_1 \end{bmatrix} + \begin{bmatrix} \zeta_1 \\ \zeta_2 \\ \zeta_3 \end{bmatrix} \end{aligned} \tag{22.6}$$

Decomposition of Effects

Decomposition of the effects of causal variables involves differentiation of the effects in terms of the mediated, or indirect effects, and the direct effect (Alwin & Hauser, 1975; Duncan, Featherman, & Duncan, 1972; Graff & Schmidt, 1982). The direct effect is the influence of the causal variables on the other variables involving "a chain of length one" (Sobel, 1987) in the sequence of causal relations. The matrices specifying the relations between two variables in the above equations provide the direct effects in a model. The $\mathbf{\Gamma}$ matrix contains the direct effects of the exogenous independent variables (ξ) on the mediators and the dependent variables (η's) and the $\mathbf{B}$ matrix contains the direct effects of the mediators on the dependent variables. The indirect effect is the effect of one variable on another variable that is intervened by at least one additional variable in the "chains of length r ($r \geq 2$)" (Sobel, 1987) causal relations. The total indirect effect is the sum of the specific indirect effects in a model, and the total effect is the sum of the total indirect effect and the direct effect (Bollen, 1987; Fox, 1980).

To illustrate how the total effects, total indirect effects, and direct effects are obtained using matrix form, let $\mathbf{T}_{\eta\eta}$ denote the matrix for the total effects of endogenous variables on endogenous variables and let $\mathbf{T}_{\eta\xi}$ denote the matrix for the total effects of exogenous variables on endogenous variables. In recursive models where $\mathbf{B}$ is a lower triangular matrix, $\mathbf{T}_{\eta\eta}$ is defined as follows (Brown, 1997; Fox, 1980):

$$\mathbf{T}_{\eta\eta} = \sum_{k=1}^{\infty} \mathbf{B}^k \tag{22.7}$$

where k is equal to powers of the direct effects between the endogenous variables in the model. For the recursive model with q endogenous variables, it is not possible to have chains whose length exceed the number of the endogenous variables, i.e., effect of length of q or higher. Thus, $\mathbf{B}^{\mathbf{q}}$ or higher-powered matrix is 0, resulting in Equation 22.8:

$$\mathbf{T}_{\eta\eta} = \mathbf{B} + \mathbf{B}^2 + \mathbf{B}^3 + \ \ldots \ + \mathbf{B}^{q-1} \tag{22.8}$$

For example, in Figure 22.2 (q = 3), the total effects of the endogenous variables on the other endogenous variables can be obtained as Equation 22.9:

$$\begin{aligned} \mathbf{T}_{\eta\eta} = \mathbf{B} + \mathbf{B}^2 &= \begin{bmatrix} 0 & 0 & 0 \\ \beta_{21} & 0 & 0 \\ \beta_{31} & \beta_{32} & 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ \beta_{21}\beta_{32} & 0 & 0 \end{bmatrix} \\ &= \begin{bmatrix} 0 & 0 & 0 \\ \beta_{21} & 0 & 0 \\ \beta_{31} + \beta_{21}\beta_{32} & \beta_{32} & 0 \end{bmatrix} \end{aligned} \tag{22.9}$$

Using the actual estimates in Figure 22.2, Equation 22.9 can be written as follows:

$$\mathbf{T}_{\eta\eta} = \mathbf{B} + \mathbf{B}^2 = \begin{bmatrix} 0 & 0 & 0 \\ \beta_{21} & 0 & 0 \\ \beta_{31} + \beta_{21}\beta_{32} & \beta_{32} & 0 \end{bmatrix}$$
$$= \begin{bmatrix} 0 & 0 & 0 \\ 0.164 & 0 & 0 \\ 0.102 + 0.164 \times 0.259 & 0.259 & 0 \end{bmatrix} \quad (22.10)$$
$$= \begin{bmatrix} 0 & 0 & 0 \\ 0.164 & 0 & 0 \\ 0.144 & 0.259 & 0 \end{bmatrix}$$

The total effect of η_1 on η_2 equals $\beta_{21} = 0.164$, and the total effect of η_1 on η_3 equals $\beta_{31} + \beta_{21}\beta_{32} = 0.144$.

To obtain the total effects of the exogenous variables on the endogenous variables, the $\boldsymbol{\Gamma}$ matrix can be incorporated as follows:

$$\mathbf{T}_{\eta\xi} = \boldsymbol{\Gamma} + \mathbf{B}\boldsymbol{\Gamma} + \mathbf{B}^2\boldsymbol{\Gamma} + \mathbf{B}^3\boldsymbol{\Gamma} + \ \ldots$$
$$+\mathbf{B}^{q-1}\boldsymbol{\Gamma} = (\mathbf{I} + \mathbf{B} + \mathbf{B}^2 + \mathbf{B}^3 + \ \ldots \ + \mathbf{B}^{q-1})\boldsymbol{\Gamma} \quad (22.11)$$

where $\mathbf{I}$ is an identity matrix of order of q. For the model in Figure 22.2, the total effects of the exogenous variable on the endogenous variables can be obtained as Equation 22.12:

$$\mathbf{T}_{\eta\xi} = (\mathbf{I} + \mathbf{B} + \mathbf{B}^2)\boldsymbol{\Gamma}$$
$$= \left(\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 \\ \beta_{21} & 0 & 0 \\ \beta_{31} & \beta_{32} & 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ \beta_{21}\beta_{32} & 0 & 0 \end{bmatrix} \right) \begin{bmatrix} \gamma_{11} \\ \gamma_{21} \\ \gamma_{31} \end{bmatrix} \quad (22.12)$$
$$= \begin{bmatrix} \gamma_{11} \\ \gamma_{11}\beta_{21} + \gamma_{21} \\ \gamma_{11}\beta_{31} + \gamma_{11}\beta_{21}\beta_{32} + \gamma_{21}\beta_{32} + \gamma_{31} \end{bmatrix}$$

Using the path coefficients between the latent variables in Figure 22.2, we can obtain the total effects of the exogenous variable on the endogenous variables as follows:

$$\mathbf{T}_{\eta\xi} = (\mathbf{I} + \mathbf{B} + \mathbf{B}^2)\boldsymbol{\Gamma} = \begin{bmatrix} \gamma_{11} \\ \gamma_{11}\beta_{21} + \gamma_{21} \\ \gamma_{11}\beta_{31} + \gamma_{11}\beta_{21}\beta_{32} + \gamma_{21}\beta_{32} + \gamma_{31} \end{bmatrix}$$
$$= \begin{bmatrix} 0.298 \\ 0.298 \times 0.164 + 0.634 \\ 0.298 \times 0.102 + 0.298 \times 0.164 \times 0.259 + 0.634 \\ \times 0.259 + 0.024 \end{bmatrix}$$
$$= \begin{bmatrix} 0.298 \\ 0.683 \\ 0.231 \end{bmatrix} \quad (22.13)$$

The total effects in Equations 22.9 and 22.11 can be simplified as follows (Bollen, 1987):

$$\mathbf{T}_{\eta\eta} = (\mathbf{I} - \mathbf{B})^{-1} - \mathbf{I} \quad (22.14)$$

$$\mathbf{T}_{\eta\xi} = (\mathbf{I} - \mathbf{B})^{-1}\boldsymbol{\Gamma} \quad (22.15)$$

Because the total indirect effects in the recursive models are obtained by subtracting the direct effects from the total effects, the total indirect effects of the endogenous variables on the endogenous variables may be calculated as follows:

$$\mathbf{I}_{\eta\eta} = \mathbf{T}_{\eta\eta} - \mathbf{B} = (\mathbf{I} - \mathbf{B})^{-1} - \mathbf{I} - \mathbf{B} \quad (22.16)$$

Applying Equation 22.16 to Figure 22.2, the total indirect effects of the endogenous variables are

$$\mathbf{I}_{\eta\eta} = \mathbf{T}_{\eta\eta} - \mathbf{B} = \begin{bmatrix} 0 & 0 & 0 \\ \beta_{21} & 0 & 0 \\ \beta_{31} + \beta_{21}\beta_{31} & \beta_{32} & 0 \end{bmatrix} - \begin{bmatrix} 0 & 0 & 0 \\ \beta_{21} & 0 & 0 \\ \beta_{31} & \beta_{32} & 0 \end{bmatrix}$$
$$= \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ \beta_{21}\beta_{31} & 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0.042 & 0 & 0 \end{bmatrix} \quad (22.17)$$

Similarly, the total indirect effects of the exogenous variables on the endogenous variables can be obtained as follows:

$$\mathbf{I}_{\eta\xi} = \mathbf{T}_{\eta\xi} - \boldsymbol{\Gamma} = (\mathbf{I} - \mathbf{B})^{-1}\boldsymbol{\Gamma} - \boldsymbol{\Gamma} \quad (22.18)$$

Applying Equation 22.18 to Figure 22.2, the total indirect effects of the exogenous variable are

$$\begin{aligned}\mathbf{I}_{\eta\xi} &= \mathbf{T}_{\eta\xi} - \boldsymbol{\Gamma} \\ &= \begin{bmatrix} \gamma_{11} \\ \gamma_{11}\beta_{21} + \gamma_{21} \\ \gamma_{11}\beta_{31} + \gamma_{11}\beta_{21}\beta_{32} + \gamma_{21}\beta_{32} + \gamma_{31} \end{bmatrix} - \begin{bmatrix} \gamma_{11} \\ \gamma_{21} \\ \gamma_{31} \end{bmatrix} \\ &= \begin{bmatrix} 0 \\ \gamma_{11}\beta_{21} \\ \gamma_{11}\beta_{31} + \gamma_{11}\beta_{21}\beta_{32} + \gamma_{21}\beta_{32} \end{bmatrix} = \begin{bmatrix} 0 \\ 0.049 \\ 0.207 \end{bmatrix}\end{aligned} \quad (22.19)$$

For example, the total indirect effect of ξ_1 on η_3 is $\gamma_{11}\beta_{31} + \gamma_{11}\beta_{21}\beta_{32} + \gamma_{21}\beta_{32} = 0.207$. The total indirect effect of ξ_1 on η_3 is further decomposed into three specific indirect effects, each of which is obtained by the product of path coefficients that are involved in the mediational pathways: (1) the indirect effect via η_1 only, i.e., $\gamma_{11}\beta_{31} = 0.030$; (2) the indirect effect via η_1 and η_2 (i.e., $\gamma_{11}\beta_{21}\beta_{32} = 0.013$); and (3) the indirect effect via η_2 only, (i.e., $\gamma_{21}\beta_{32} = 0.164$).

Standard Errors of Indirect Effects

In order to conduct significance tests or to construct confidence intervals for the indirect effects, standard errors of the indirect effects are needed. One of the most widely used methods for estimating standard errors of indirect effects is the method derived by Sobel (1982, 1986) based on first derivatives using the multivariate delta method (Bishop, Fienberg, & Holland, 1975; Folmer, 1981). According to the multivariate delta method, the asymptotic variance of a function of random variables, such as the product of path coefficients, can be obtained by pre- and postmultiplying the covariance matrix of the random variables by a vector of first partial derivatives of the random variable function. For example, let $\theta = [\theta_1, \theta_2, \ldots, \theta_u]'$ be a vector of direct effects, $f(\theta)$ be a function of direct effects, $\Sigma(\theta)$ be the asymptotic covariance matrix of the direct effects, and $f^1(\theta)$ be the first derivative of $f(\theta)$ with respect to θ, that is, $[\partial f/\partial\theta_1, \partial f/\partial\theta_2, \ldots, \partial f/\partial\theta_u]'$. Then, the variance of the function of random variables is as follows (Sobel, 1982, 1987):

$$V(f(\theta)) = [f^1(\theta)]'\Sigma(\theta)f^1(\theta) \quad (22.20)$$

Provided $f(\theta)$ is continuously differentiable and $\Sigma(\theta)$ is a continuous function of θ, the quantity in Equation 22.20 can be estimated using the sample estimates as follows:

$$V(f(\hat{\theta})) = [f^1(\hat{\theta})]'S(\hat{\theta})f^1(\hat{\theta}) \quad (22.21)$$

where $S(\hat{\theta})$ is the sample estimate of $\Sigma(\theta)$. Then, the square root of the term in Equation 22.21 can be used for significance testing and confidence interval estimation of $f(\theta)$.

Applying the delta method described above to the specific indirect effect of ξ_1 on η_2 via η_1 ($\xi_1 \rightarrow \eta_2 \rightarrow \eta_3$) in Figure 22.2, $f(\hat{\theta}) = \hat{\gamma}_{21}\hat{\beta}_{32}$, where $\hat{\gamma}_{21}$ and $\hat{\beta}_{32}$ are sample estimates of the two path coefficients, and $[f^1(\hat{\theta})]' = [\partial f/\partial\hat{\gamma}_{21} \quad \partial f/\partial\hat{\beta}_{32}] = [\hat{\beta}_{32} \quad \hat{\gamma}_{21}]$. With the sample estimates of $S(\hat{\theta})$, the estimated variance of the indirect effect $\hat{\gamma}_{11}\hat{\beta}_{31}$ is obtained as follows:

$$\begin{aligned}Var(\hat{\gamma}_{21}\hat{\beta}_{32}) &= [\hat{\beta}_{32} \quad \hat{\gamma}_{21}] \begin{bmatrix} s^2_{\gamma_{21}} & s_{\gamma_{21}\beta_{32}} \\ s_{\beta_{32}\gamma_{21}} & s^2_{\beta_{32}} \end{bmatrix} \begin{bmatrix} \hat{\beta}_{32} \\ \hat{\gamma}_{21} \end{bmatrix} \\ &= \hat{\gamma}^2_{21}s^2_{\beta_{32}} + \hat{\beta}^2_{32}s^2_{\gamma_{21}} + 2\hat{\gamma}_{21}\hat{\beta}_{32}s_{\gamma_{21}\beta_{32}}\end{aligned} \quad (22.22)$$

where $s^2_{\beta_{32}}$, $s^2_{\gamma_{21}}$, and $s_{\gamma_{21}\beta_{32}}$ are the maximum likelihood estimates of the variances and covariance of the relevant path coefficients. Using the estimates in Figure 22.2, the variance in Equation 22.22 can be calculated as follows:

$$\begin{aligned}Var(\hat{\gamma}_{21}\hat{\beta}_{32}) &= [\hat{\beta}_{32} \quad \hat{\gamma}_{21}] \begin{bmatrix} s^2_{\gamma_{21}} & s_{\gamma_{21}\beta_{32}} \\ s_{\beta_{32}\gamma_{21}} & s^2_{\beta_{32}} \end{bmatrix} \begin{bmatrix} \hat{\beta}_{32} \\ \hat{\gamma}_{21} \end{bmatrix} \\ &= [0.259 \quad 0.634] \begin{bmatrix} 0.002 & 0 \\ 0 & 0.002 \end{bmatrix} \begin{bmatrix} 0.259 \\ 0.634 \end{bmatrix} \\ &= 0.634^2 \times 0.002 + 0.259^2 \times 0.002 + 2 \times 0.634 \times 0 \\ &= 0.001\end{aligned} \quad (22.23)$$

In the single mediator model, the covariance between coefficients is equal to zero, resulting in the simplified formula, $Var(\hat{\gamma}_{21}\hat{\beta}_{32}) = \hat{\gamma}^2_{21}s^2_{\beta_{32}} + \hat{\beta}^2_{32}s^2_{\gamma_{21}}$. In many models, however, the covariances between path coefficients may be nonzero and more accurate standard errors can be obtained by including the covariance terms. In the supplementary materials, we show how the rules above can be applied to derive the standard error for multiple indirect effects.

Other researchers have suggested different methods to estimate the standard error of indirect effects (MacKinnon & Dwyer, 1993). For example, Allison (1995) used a reduced form parameterization for indirect effect models with no latent variables estimated with ordinary least squares estimation and

derived the same standard error formula as Sobel (1982, 1986). The formula derived with second-order Taylor series and exact variance (Goodman, 1960; Mood, Graybill, & Boes, 1974) includes the product of the variances of the path coefficients, instead of the covariance between the path coefficients, resulting in $Var(\hat{\gamma}_{21}\hat{\beta}_{32}) = \hat{\gamma}_{21}^2 s_{\beta_{32}}^2 + \hat{\beta}_{32}^2 s_{\gamma_{21}}^2 + s_{\beta_{32}}^2 s_{\gamma_{21}}^2$. In the formula for the unbiased variance estimator shown by Goodman (1960), the product of the variances of the path coefficients is subtracted rather than added, $Var(\hat{\gamma}_{21}\hat{\beta}_{32}) = \hat{\gamma}_{21}^2 s_{\beta_{32}}^2 + \hat{\beta}_{32}^2 s_{\gamma_{21}}^2 - s_{\beta_{32}}^2 s_{\gamma_{21}}^2$. Despite the differences in the formulas, the estimates of the standard error of the indirect effect are very close to each other. However, simulation studies show that the estimator based on the first-order multivariate delta method (Sobel, 1982) produces the least biased estimates (MacKinnon & Dwyer, 1993; MacKinnon, Warsi, & Dwyer, 1995; MacKinnon, Lockwood, Hoffman, West, & Sheets, 2002). The first-order multivariate delta method estimator is implemented in most SEM software programs including EQS (Bentler, 1997), LISREL (Jöreskog & Sörbom, 2006), Amos (Arbuckle & Wothke, 1999), lavaan in R (Rosseel, 2012), and Mplus (Muthén & Muthén, 1998/2014). More detailed information about different estimators of the standard error of the indirect effect can be found in MacKinnon and colleagues (2002).

Confidence Intervals of Indirect Effects

The application of the first-order multivariate delta method described above requires three conditions: (1) the parameters in $\hat{\theta}$ are consistent and normally distributed estimator of θ in large samples; (2) the indirect effect, $f(\hat{\theta})$, is continuously differentiable with respect to θ; and (3) $\Sigma(\hat{\theta})$ is a continuous function of $\hat{\theta}$. Under these conditions, the ratio of the indirect effect to its standard error is approximately normally distributed in large samples with mean of 0 and variance of 1. Thus, using the standard normal distribution, the confidence limits for the indirect effect $f(\hat{\theta})$ is obtained as follows:

$$f(\hat{\theta}) \pm z_{Type\ 1\ error}\sqrt{[f^1(\hat{\theta})]' S(\hat{\theta}) f^1(\hat{\theta}))} \quad (22.24)$$

Applying this rationale to the specific indirect effect ξ_1 on η_2 via η_1 ($\xi_1 \rightarrow \eta_2 \rightarrow \eta_3$) in Figure 22.2, the indirect effect $\hat{\beta}_{32}\hat{\gamma}_{21}$ is normally distributed with the variance of $\hat{\gamma}_{21}^2 s_{\beta_{32}}^2 + \hat{\beta}_{32}^2 s_{\gamma_{21}}^2 + 2\hat{\gamma}_{21}\hat{\beta}_{32} s_{\gamma_{21}\beta_{32}}$. The 95% confidence intervals for $\hat{\beta}_{32}\hat{\gamma}_{21}$ are then calculated as follows:

$$\begin{aligned} &\hat{\beta}_{32}\hat{\gamma}_{21} \pm 1.965\sqrt{\hat{\gamma}_{21}^2 s_{\beta_{32}}^2 + \hat{\beta}_{32}^2 s_{\gamma_{21}}^2 + 2\hat{\gamma}_{21}\hat{\beta}_{32} s_{\gamma_{21}\beta_{32}}} \quad (22.25)\\ &= 0.164 \pm 1.965\sqrt{0.001} \end{aligned}$$

The resulting lower confidence limit is 0.102 and the upper confidence limit is 0.226. Because the sample interval does not include 0, the indirect effect is considered statistically significant.

Because the standard error of the indirect effect described earlier is based on the normal theory of the product of path coefficients, the confidence intervals obtained from Equation 22.24 are assumed to be symmetrical, with the upper and the lower limits at an equal distance from the estimated indirect effect. Simulation studies, however, demonstrate that the distributions of indirect effects based on the product of coefficients are often non-normal and confidence intervals based on Equation 22.24 tend to fall to the left of the true value of the indirect effect more often than to the right (MacKinnon et al., 1995, 2004; Stone & Sobel, 1990), although, with increased sample sizes, the proportions of times that confidence intervals fall on either side of the true value become similar. The imbalance of confidence intervals and the non-normality of the distributions of indirect effects have been also observed when bootstrapping methods are used (Bollen & Stine, 1990; MacKinnon et al., 2004).

Alternative methods have been developed to consider the non-normality of distributions of indirect effects. Confidence intervals of indirect effects can be constructed using the distribution of the product of normally distributed random variables (Craig, 1936; Meeker, Cornwell, & Aroian, 1981; Springer & Thompson, 1966), which are usually asymmetrical, unlike those calculated using Equation 22.24 (MacKinnon et al., 2002, 2004; MacKinnon, Fritz, Williams, & Lockwood, 2007). Software programs such as PRODCLIN (MacKinnon et al., 2007) and the RMediation package in R (Tofighi & MacKinnon, 2011, 2016) are available for computing these confidence intervals of indirect effects. Furthermore, likelihood-based confidence intervals (Cheung, 2007, 2009) can capture asymmetrical distributions of indirect effects based on log-likelihood functions (for details, see Meeker & Escobar, 1995; Neale & Miller, 1997). The Bayesian approach to mediation analysis (Enders, Fairchild, & MacKinnon, 2013; Miočević, MacKinnon, & Levy, 2017; Yuan & MacKinnon, 2009) can also accommodate the non-normal distribution of the indirect effect by empirically approximating the posterior distribution

of the indirect effect Last, resampling methods such as the bootstrap (Efron & Tibshirani, 1993), permutation (Edgington, 1995), randomization (Edgington, 1995), and jackknife (Mosteller & Tukey, 1977) can be used to build confidence intervals of indirect effects (MacKinnon et al., 2004; Taylor, MacKinnon, & Tein, 2008). In these methods, a large number of resamples are taken from the original observed sample, the indirect effect is computed in each of the resamples, and the creation of the empirical distribution of the indirect effect is based on the indirect effects computed across the resamples. The mean and the standard deviation of this distribution are the bootstrap estimate of the indirect effect and standard error of the indirect effect, respectively. Construction of the confidence intervals is then based on the empirical distributions, for example, by obtaining the 2.5th and 97.5th percentiles. There are different bootstrap methods, including the percentile bootstrap, the bias-corrected bootstrap, and the bias-corrected and accelerated bootstrap confidence intervals. Bootstrap procedures are now available in SEM programs such as Amos, Mplus, lavaan in R, and EQS. SPSS and SAS codes for using the bootstrap method for simpler models are also available (Preacher & Hayes, 2004, 2008). Simulation studies examining statistical performances of these methods demonstrated that the methods that do not assume normal sampling distributions of indirect effects performed better than normal theory-based methods in terms of accuracy of estimates, power, and coverage (Biesanz, Falk, & Savalei, 2010; Cheung, 2009; Falk & Biesanz, 2015; MacKinnon et al., 2002, 2004; Valente, Gonzalez, Miočević, & MacKinnon, 2016). Bayesian credibility intervals for indirect effects have shown power comparable to the distribution of the product and bootstrap (i.e., percentile and bias-corrected bootstrap) confidence intervals, and the gain from incorporating prior information was generally greater when the sample size was small or moderate (i.e., smaller than 200; Miočević, MacKinnon, & Levy, 2017; Yuan & MacKinnon, 2009).

ADVANCES IN MEDIATION ANALYSIS

Recent advances have extended the concept of mediation analysis to more complex models, such as multilevel models or longitudinal models. We discuss multilevel mediation (Krull & MacKinnon, 1999, 2001; Preacher, Zyphur, & Zhang, 2010) and longitudinal mediation (Cheong, 2011; Cheong, MacKinnon, & Khoo, 2003) in the supplementary materials. Here we focus on the causal interpretation of the mediated effect and the measurement of mediators.

Causal Inference

One way to view the use of SEM is that it provides a way to consider all variables relevant to a phenomenon under study. This contrasts with examining bivariate relations separately, which has been criticized for excluding the omitted variables that would be relevant for the bivariate relation. From this perspective, SEM was developed to provide a solution to the omitted variables problem for causal inference by including as many relevant variables as possible for a phenomenon under study. Nevertheless, causal inference in SEM, including estimation of indirect effects, has been a controversial topic since it was first introduced. As mentioned earlier, Wright (1920, 1921) argued that path analysis was not a method to infer causation; it is a method to quantify assumed causal relations including chains of mediation.

While much work on causal inference has focused on the relation between two variables *X* and *Y,* demonstrating the importance of randomization for causal inference of *X* to *Y,* as well as other assumptions (Angrist, Imbens, & Rubin, 1996; Frangakis & Rubin, 2002; Holland, 1988a, 1988b), more recent work in this area has focused on causal inference for mediating mechanisms. Benefits of causal inference for mediation are the clarification of causal effects with more precise mathematical definition of causal relations, including criteria based on actual and counterfactual conditions, along with considerations of the limitations and strengths of different types of evidence for mediation (Ten Have et al., 2007; VanderWeele, 2008; VanderWeele & Robins, 2008; VanderWeele & Vansteelandt, 2009).

Recently, the identification of causal mediation effects has received a lot of attention (Imai, Keele, & Tingley, 2010; Pearl, 2001, 2009; Robins & Greenland, 1992; VanderWeele & Vansteelandt, 2009). Imai and colleagues (2010) and Pearl (2014) describe theoretical differences between the various no unmeasured confounder assumptions used for identification. The most referenced no unmeasured confounder assumptions to nonparametrically identify causal mediation effects are as follows:

1. No unmeasured confounders of the *X*–*Y* relation conditional on covariates.

2. No unmeasured confounders of the *M*–*Y* relation conditional on *X* and covariates.
3. No unmeasured confounders of the *X*–*M* relation conditional on covariates.
4. No confounders of the *M*–*Y* relation affected by *X* conditional on covariates.

Assumptions 1 and 3 are satisfied in expectation when *X* represents random assignment to an experimental condition. Assumption 2 is not satisfied when *X* represents random assignment to an experimental condition because individuals still self-select their mediator value given their observed treatment and covariate values. Assumption 4 is referred to as the *recanting witness assumption* (Avin, Shpitser, & Pearl, 2005; Tchetgen & VanderWeele, 2014) because identification of the natural direct and indirect effects requires that a confounder of the *M*–*Y* relation that is affected by *X* is held constant at both the control group and treatment group level simultaneously; therefore, the posttreatment confounder variable "has it both ways." Another way to conceptualize Assumption 4 is that there are no additional mediators that were affected by the intervention that in turn affect both the focal mediator and the outcome resulting in a serial multiple mediator model.

Considering counterfactual conditions—conditions that a research participant had not served in as well as the condition in which they had served—enabled a mathematical formulation of causal inference (Pearl, 2001, 2012). In total, there are six causal mediation effects, all of which are defined as the difference between potential outcomes (counterfactuals): (1) total natural indirect effect, (2) pure natural indirect effect, (3) total natural direct effect, (4) pure natural direct effect, (5) controlled direct effect, and (6) total effect.

Suppose that *X* represents assignment to a randomized condition for which $X = 0$ represents assignment to the control group and $X = 1$ represents assignment to the treatment group. The total natural indirect effect (TNIE) is the effect of *X* on *Y* through *M,* when holding the treatment constant at the treatment-group level $X = 1$; TNIE = $E[Y_i (1, M_i (1)) - Y_i (1, M_i 0))]$. The pure natural indirect effect (PNIE) is the effect of *X* on *Y* through *M,* when holding the treatment constant at the control-group level $X = 0$; PNIE = $E[Y_i (0, M_i (1)) - Y_i (0, M_i 0))]$. The total natural direct effect (TNDE) is the effect of *X* on *Y,* while fixing each individual's mediator to the value that would naturally have been observed had the individual been in the treatment group (i.e., $M_i (1)$); TNDE = $E[Y_i (1, M_i (1)) - Y_i (0, M_i (1))]$. The pure natural direct effect (PNDE) is the effect of *X* on *Y,* while fixing each individual's mediator to the value that would naturally have been observed had the individual been in the control group (i.e., $M_i (0)$); PNDE = $E[Y_i (1, M_i (0)) - Y_i (0, M_i (0))]$. The controlled direct effect (CDE) is the effect of *X* on *Y,* while fixing the mediator for all individuals at a predetermined value *m*; CDE = $E[Y_i (1, m) - Y_i (0, m)]$. The total effect (TE) is the effect of *X* on *Y* and equals the sum of the TNDE and the PNIE or, equivalently, the sum of the PNDE and the TNIE. Additional decompositions of the total effect have been described by VanderWeele (2014). Because the potential outcomes framework provides nonparametric causal mediation effect definitions, they can be directly applied to a wide range of models that may include nonadditive or nonlinear effects.

MacKinnon, Valente, and Gonzalez (2020) demonstrated in linear models with continuous *M,* continuous *Y,* and *X*–*M* interaction a correspondence between causal mediation effects and simple regression effects. The term "simple" refers to the relation between two variables (e.g., *M* and *Y*) at one level of the independent variable *X* (borrowing the term from analogous simple effects and simple slope tests in regression and analysis of variance; Aiken & West, 1991). In the presence of an *X*–*M* interaction, main effects of *X* on *Y* and *M* on *Y* provide an incomplete picture of the relations in the mediation model. The implication of a statistically significant *X*–*M* interaction is that the mediated effect is moderated by *X,* and the direct effect is moderated by *M.* The PNIE is analogous to the simple indirect effect in the control group; the TNIE is analogous to the simple indirect effect in the treatment group; the PNDE is analogous to the simple direct effect in the control group; and the TNDE is analogous to the simple direct effect in the treatment group. The difference between the total and pure effects (or the simple effects in the treatment and control groups) is referred to as the "mediated interaction." When there is no *X*–*M* interaction or other nonlinearities (e.g., a binary outcome), the TNDE = PNDE, and the TNIE = PNIE. For an overview of causal mediation effects with binary outcomes, see VanderWeele and Vansteelandt (2010) and Rijnhart, Valente, MacKinnon, Twisk, and Heymans, (2021).

Our example uses data from the ATLAS program. The exogenous variable (*X*) was the group membership (i.e., treatment vs. control), and strength training self-efficacy was the outcome (*Y*). The mediator (*M*) was knowledge of the effects of anabolic androgenic steroids (AAS). The sample size was 1,191 after listwise

deletion of the original 1,506 observations. The causal mediation effects were estimated using the Valeri and VanderWeele (2013, 2015) SAS macro accompanied by 95% percentile bootstrap confidence intervals. The CDE was evaluated at the mediator value of zero, which represented the mean of the mediator because it was mean-centered. For a review of available software for estimating causal mediation effects, see Valente, Rijnhart, Smyth, Muniz, and MacKinnon (2020). The PNIE was equal to 0.203 (lower confidence limit [LCL] = 0.148, upper confidence limit [UCL] = 0.268), the TNIE was equal to 0.177 (LCL = 0.116, UCL = 0.250), the PNDE was equal to 0.125 (LCL = –0.006, 0.251), the TNDE was equal to 0.099 (LCL = –0.020, UCL = 0.217), the CDE was equal to 0.114 (LCL = –0.001, UCL = 0.231), and the TE was equal to 0.302 (LCL = 0.178, 0.429). The indirect effects of ATLAS on strength training self-efficacy through its effect on knowledge of AAS were statistically significant, while the direct effects of ATLAS on strength training self-efficacy were not. The *X*–*M* interaction was equal to –0.010 (LCL = –0.035, UCL = 0.015) and the mediated interaction was equal to –0.027 (LCL = –0.101, UCL = 0.039), neither of which were statistically significant for this example.

Posttreatment Confounding

Identification of causal mediation effects become more challenging in the presence of posttreatment confounding (i.e., when Assumption 4, recanting witness, is violated). Figure 22.3 presents a directed acyclic graph (DAG) of a single-mediator model with a posttreatment confounder, L. Identification of the CDE in the presence of posttreatment confounding requires fewer assumptions than identification of natural direct and indirect effects (Moerkerke, Loeys, & Vansteelandt, 2015; Pearl, 2001; VanderWeele & Vansteelandt, 2009). An important point is that the direct effects are defined as the effect of *X* on *Y not through M*; therefore, in the presence of posttreatment confounding L, the direct effects are a combination of the pathways $X \rightarrow Y$ plus $X \rightarrow \mathrm{L} \rightarrow Y$. To estimate the CDE in standard SEM, an estimate of the path $X \rightarrow \mathrm{L} \rightarrow Y$ needs to be incorporated into the estimate of the traditional direct effect estimate $\mathrm{X} \rightarrow \mathrm{Y}$. Standard SEM implicitly makes an unnecessary no-unmeasured-confounder assumption: no unmeasured confounders of the posttreatment confounder (L) and the outcome (*Y*). This additional no-unmeasured-confounder assumption is implied by standard uncorrelated residuals assumptions in SEM (De Stavola, Daniel, Ploubidis, & Micali, 2015; Moerkerke et al., 2015). Therefore, SEM makes stricter causal (and parametric) assumptions than are necessary to identify the CDE in the presence of posttreatment confounding. While the total and pure direct and indirect effects would not be nonparametrically identified in this instance, they would be identified if parametric models for all relations are assumed (Pearl, 2014). Additionally, natural direct and indirect effects may be identified in the presence of posttreatment confounding under various sets of assumptions that are less restrictive than assuming correctly specified parametric models (Didelez, Dawid, & Geneletti, 2006; Petersen, Sinisi, van der Laan, 2006; Tchetgen & VanderWeele, 2014) or by treating the posttreatment confounder as an additional mediator variable (e.g., Daniel, De Stavola, Cousens, & Vansteelandt, 2015; Imai & Yamamoto, 2013; Steen, Loeys, Moerkerke, & Vansteelandt, 2017; VanderWeele, Vansteelandt, & Robins, 2014).

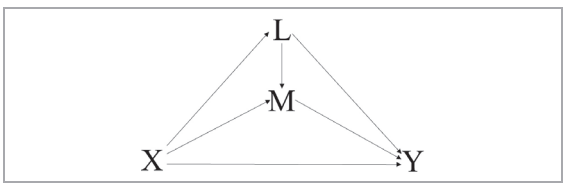

FIGURE 22.3. A directed acyclic graph (DAG) of a single mediator model with posttreatment confounder.

Three methods to handle posttreatment confounding in mediation models are G-computation, sequential G-estimation, and inverse propensity weighting (IPW) (Coffman & Zhong, 2012; Kisbu-Sakarya, MacKinnon, Valente, & Çetinkaya, 2020; Robins, 1986; Ten Have et al., 2007; Vansteelandt, 2009). G-computation relies on fitting a regression or SEM for the mediator, the outcome, and the posttreatment confounder and computing the nested potential outcomes—similar to the nested potential outcomes described earlier, but with the addition of the posttreatment confounder effects on the mediator and the outcome (Robins, 1986). Sequential G-estimation first removes the effect of the mediator on the outcome, then estimates the effect of *X* on *Y* that is remaining (Goetgeluk, Vansteelandt, & Goetghebeur, 2008; Joffe & Greene, 2009; Vansteelandt, 2009). The rationale is that once the effect of the mediator on the outcome is removed, only a direct effect

remains (Goetgeluk et al., 2008; Vansteelandt, 2009). Assuming no *X*–*M* interaction, the mediated effect can be estimated as the difference between the total effect and the G-estimate of the direct effect. Significance testing can be performed by using percentile bootstrapping to create confidence intervals for the mediated effect. IPW is an application of propensity scores, which have been described in detail for causal treatment effect estimation (Imbens & Rubin, 2015; Rosenbaum & Rubin, 1983). VanderWeele (2009) described IPW of marginal structural models (i.e., weighted regression equations) for mediation analysis. The weights for IPW are used to adjust for the influence of confounders in a weighted regression analysis to estimate mediated and direct effects that are free of the confounding effects of measured confounders. For continuous variables, it is important to estimate stabilized inverse propensity weights (Robins, Hernán, & Brumback, 2000). Stabilized weights consist of a ratio of two propensity scores and ensure finite sampling variability of the IPW estimator (Naimi, Moodie, Auger, & Kaufman, 2014; Robins, 2000; Robins et al., 2000). The size of the weights is influenced by the distributional form of the mediator (e.g., continuous mediator; Robins et al., 2000) and by the strength of the predictors of the continuous mediator. Stronger predictors lead to extreme weights more often than do weaker predictors (Goetgeluk et al., 2008; Robins et al., 2000; Valente, MacKinnon, & Mazza, 2020; Vansteelandt, 2009) which increases the sampling variability of the IPW estimator (Cole & Hernán, 2008; Hernán & Robins, 2006). To reduce the sampling variability of the IPW estimator, Cole and Hernán (2008) suggested truncating extremely large weights and extremely small weights by setting them equal to the 99th, 95th, or 90th percentile of the distribution of weights or the 1st, 5th, or 10th percentile of the distribution of weights, respectively.

Sensitivity Analyses

Sensitivity analyses offer a way to assess the robustness of mediated effect estimates in the presence of unmeasured confounding. This is particularly useful because the no-unmeasured-confounder assumptions are untestable assumptions. Thus, sensitivity analyses provide information about how the observed mediated effect estimate may be affected by plausible unmeasured confounders (Cox, Kisbu-Sakarya, Miočević, & MacKinnon, 2013).

The most versatile method for sensitivity analyses is the correlated-residuals method (Imai, Keele, & Yamamoto, 2010; Muthén, Muthén, & Asparouhov, 2017). This method starts by assuming that unmeasured confounder of the *M*–*Y* relation would result in a correlation between the residuals of the mediator and outcome equations. When the correlation between these residuals is exactly equal to zero, the observed mediated effect is identified. If the mediator–outcome relation is constrained to zero, a correlation between the residuals can be estimated. A sensitivity parameter is then varied to assess the impact of varying magnitudes of this confounder-induced correlation on the causal mediated effect estimate. The correlated-residuals method is also implemented in Mplus, which provides similar plots to the mediation R-package (Tingley, Yamamoto, Hirose, Keele, & Imai, 2014) with the addition of sensitivity plots for the causal direct effects. In another method, the left-out variables error (LOVE; Mauro, 1990) plots assume there is an unmeasured confounder that affects *M* and *Y*. LOVE plots vary the correlation between *U* and *M* and the correlation between U and *Y*. Researchers can then examine the magnitude of correlation between an unmeasured variable *U* and its relation with *M* and *Y* separately to determine if it is theoretically plausible to have a variable that, left out of the analysis, could bias the mediated effect estimate such that the mediated effect estimate would equal zero. Finally, VanderWeele and colleagues have extended sensitivity analyses to account for binary confounder of the *M*–*Y* relation (VanderWeele, 2010) and sensitivity analyses for binary outcomes with binary and/or continuous mediator and confounder (Ding & VanderWeele, 2016; Smith & VanderWeele, 2019).

All three sensitivity analysis methods just described assume that unmeasured confounding only affects the mediator–outcome relation and that the *X*–*M* and *X*–*Y* relations are not confounded but could be extended to the nonrandomized *X* case. These sensitivity analyses have different sensitivity parameters but work similarly. Researchers can use sensitivity analyses to determine if the mediated effect is robust to potential unmeasured confounding by showing that implausibly large values of these parameters are necessary to eliminate (i.e., bring to zero) the observed mediated effect. Whether the size of the sensitivity parameter is plausible depends on a priori information about the relation between the mediator and a potential unmeasured confounder, and the relation between the outcome and a potential unmeasured confounder. If no plausible value of the respective sensitivity parameters exists,

we can conclude that the mediated effect is robust to the influence of the potential unmeasured confounder. Additional sensitivity analyses have been proposed to probe the robustness of mediated effect estimates in the presence of both unmeasured confounding and measurement error (Fritz, Kenny, & MacKinnon, 2016; Liu & Wang, 2021).

Measurement Assumptions on the Mediator

In order to obtain a greater understanding of causal phenomena, a clear interpretation and precise measurement of the hypothesized mechanisms are needed (Gonzalez & MacKinnon, 2021). In the physical sciences, crucial discoveries were made once researchers were able to measure unobserved mediating variables. For example, the measurement of the genome could explain how parent characteristics were passed on to the offspring, or specific atomic reactions could explain how some chemical inputs lead to a chemical output (Gonzalez & MacKinnon, 2021; MacKinnon, 2008). In the social sciences, mediators are latent constructs assessed by asking respondents to answer a set of items (MacKinnon, 2008). Recent work has focused on evaluating how failing to meet psychometric assumptions about variables in the mediation model affects the conclusions drawn from mediation analyses (e.g., Fritz et al., 2016; Georgeson, Valente, & Gonzalez, 2021; Gonzalez & MacKinnon, 2018, 2021; Hoyle & Kenny, 1999; Ledgerwood & Shrout, 2011; Olivera-Aguilar, Rikoon, Gonzalez, Kisbu-Sakarya, & MacKinnon, 2018; Pek & Hoyle, 2016). Below, we discuss several psychometric assumptions that researchers make about the mediator; these conclusions could be largely extended to the other variables in the mediation model.

First, researchers often assume that the scores on the mediating variable are reliable. "Reliability" refers to the precision of the scores, or the proportion of the observed score variance attributable to variation on the assessed construct. In single-mediator models, measurement error in the mediator can attenuate the mediated effect (Hoyle & Kenny, 1999), but in more complex models, the effects of measurement error on the estimation of the mediated effect might be intractable (Gonzalez & MacKinnon, 2021) or even offset the effects of confounding on different paths of the mediation model (Fritz et al., 2016). Using a latent variable could accommodate measurement error in the mediator; this tends to be the preferred method to handle error over synthetic adjustments using the reliability estimate of the variable (Cole & Preacher, 2014). However, mediation paths involving latent variables could have larger standard errors relative to their point estimate compared to the standard errors of mediation paths involving observed scores (Gonzalez & MacKinnon, 2021; Legerwood & Shrout, 2011). In other words, researchers would have lower precision on an unbiased estimate using the model with latent variables rather than higher precision on an attenuated estimate using the model with observed variables.

Researchers also assume that the scores on the mediating construct can be ascribed a valid interpretation. "Validity" refers to the amount of evidence and theoretical rationale that supports the interpretation and use of a score (Messick, 1989). Sources of evidence that help support score interpretation are the examination of content of the measure (i.e., content validity), correlations of the score with external criteria (i.e., criterion validity), and the relations among the items that comprise the score (i.e., construct validity). In practice, researchers derive interpretations of the latent variable models fit to the items assessing the mediator, such as the number of dimensions underlying the set of items and which items load on each latent variable. Two threats to the valid interpretation of mediation scores are construct underrepresentation (e.g., the measure of the mediator does not assess all aspects of the true mediating construct) or construct-irrelevant variance (e.g., the measure of the mediator assesses more aspects than the true mediating construct; Gonzalez & MacKinnon, 2021). These threats are particularly relevant when researchers hypothesize that the mediator is a multifaceted construct assessed with a multifaceted measure. In this case, researchers could characterize the mediator differently: The mediator could be just one of the specific facets or an aggregation of them. If the true mediator is one facet of the construct but is represented in the model as an aggregation of all the facets, then the mediator has construct-irrelevant variance. If the true mediator is an aggregation of all facets of the construct but the model includes only one facet of the mediator, then the mediator has construct underrepresentation. Construct underrepresentation in the mediator would produce models that do not fully capture the causal sequence from independent variable to outcome. Construct-irrelevant variance in the mediator would lead to lower power and an attenuation of the indirect effect (Gonzalez & MacKinnon, 2021). In this case, identifying the true mediator in the causal sequence could be seen as a measurement problem in which multidimensional latent variable

models (e.g., bifactor models, higher-order factor models, or correlated-factor models; see Figure 22.4) are used to distill the mediating variable and probe indirect effects through different levels of generality of the construct (Gonzalez & MacKinnon, 2018).

Another psychometric assumption of the mediation model is that the relation between the items assessing the mediator and the mediator latent variable is the same across the independent variable (e.g., treatment and control group; Gonzalez & MacKinnon, 2021). If this assumption is violated, then the mediator may suffer from measurement bias (i.e., measurement noninvariance; Millsap, 2011), which could manifest as spurious *X*–*M* interactions (MacKinnon et al., 2020). Noninvariance in the mediator could be due to response shift effects (Oort, Visser, & Sprangers, 2009; Sprangers & Schwartz, 1999), or a change in the meaning of an individual's self-evaluation on the construct (i.e., appraisal), which could be a by-product of an intervention. Specifically, individuals who participate in an intervention could learn more about the construct compared to a control group, so the frame of reference when responding to the mediator measure differs across treatment and control respondents. In the single-mediator and two-wave mediator models, measurement bias on the mediator (specifically, on the item intercepts) could lead to high Type 1 and Type 2 error rates, along with biased indirect effect estimates depending on the direction of the measurement bias (Georgeson et al., 2021; Olivera-Aguilar et al., 2018). In summary, scores on the mediating variable should be reliable, free of bias, and interpretable because the conclusions that we make from statistical models go as far as the psychometric assumptions we can meet.

SUMMARY AND CONCLUSIONS

Mediation analysis is beneficial in that it provides information on the underlying mechanisms by which the independent variable affects the dependent variable. Especially in experimental studies such as randomized clinical trials, mediation analysis informs researchers of whether the experimental manipulation has worked as hypothesized and whether the intermediate variable is a good mediator as specified in the theory (MacKinnon & Dwyer, 1993). Such detailed information can be further used to improve intervention/prevention programs for future studies. In recent years, extensive research has been conducted on the statistical approaches to testing indirect effects. Methods for estimating indirect effects and their standard error have been developed, and numerous simulation studies have been conducted to evaluate statistical performances of

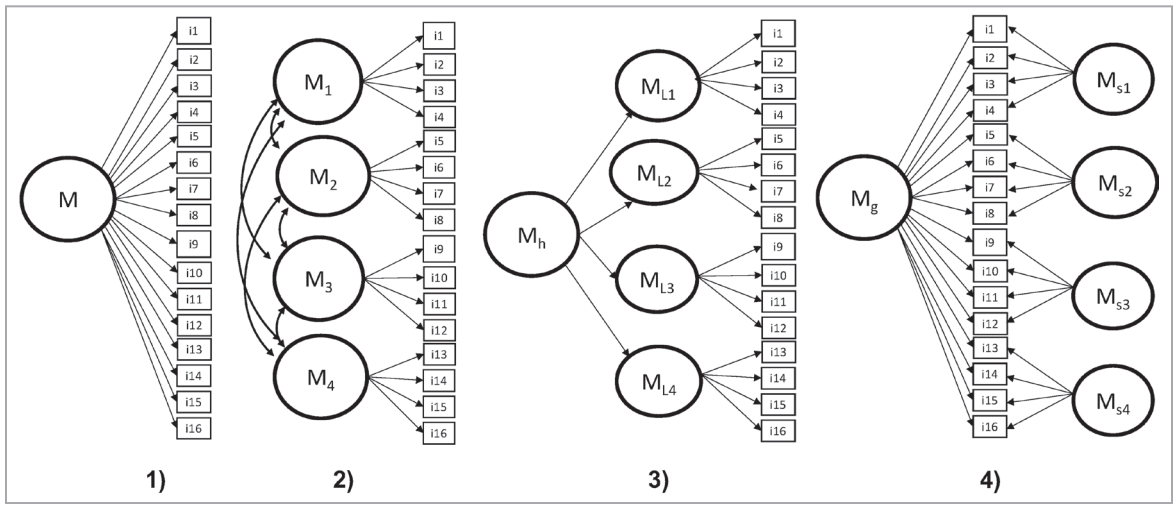

FIGURE 22.4. Latent variable models for the mediator: 1) a unidimensional model; 2) a correlated-factors model; 3) a higher-order factor model with lower-order factor model; and 4) a bifactor model with four specific factors.

these methods. Research in causal inference for mediation effects and structural equation models is currently an active field of study and has led to many recent developments in the understanding of causal mechanisms in linear and nonlinear models. Several new methods have emerged—such as G-computation, sequential G-estimation, and IPW—to handle adjustment for baseline and posttreatment confounder variables. Several methods have been developed to probe the robustness of untestable no-unmeasured-confounder assumptions via sensitivity analyses. Central to the understanding of causal mechanisms is measurement theory. Measurement theory of the mediating process involves reliable and valid measures of the mediator variable. Psychometric theory is important to ensure the mediator variable is neither too broadly nor too narrowly defined with respect to the mediation theory under investigation and to ensure any measurement bias has been eliminated. These recent developments in causal inference and measurement for mediation have accelerated our understanding of the limitations and strengths of contemporary mediation analysis methods.

REFERENCES

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

Allison, P. D. (1995). Exact variance of indirect effects in recursive linear models. *Sociological Methodology, 25*, 253–266.

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

Angrist, J. D., Imbens, G. W., & Rubin, D. B. (1996). Identification of causal effects using instrumental variables. *Journal of the American Statistical Association, 91*, 444–472.

Arbuckle, J. L., & Wothke, W. (1999). *Amos users' guide*. Chicago: Smallwaters Corporation.

Avin, C., Shpitser, I., & Pearl, J. (2005, July). Identifiability of path-specific effects. *Proceedings to the International Joint Conferences on Artificial Intelligence*, pp. 357–363.

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.

Bentler, P. M. (1980). Multivariate analysis with latent variables: Causal modeling. *Annual Review of Psychology, 31*, 419–456.

Bentler, P. M. (1997). *EQS for Windows (version 5.6)* [Computer program]. Encino, CA: Multivariate Software, Inc.

Biesanz, J. C., Falk, C. F., & Savalei, V. (2010). Assessing mediational models: Testing and interval estimation for indirect effects. *Multivariate Behavioral Research, 45*, 661–701.

Bishop, Y. M. M., Fienberg, S. E., & Holland, P. W. (1975). *Discrete multivariate analysis: Theory and practice*. Cambridge, MA: MIT Press.

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

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

Brown, R. L. (1997). Assessing specific mediational effects in complex theoretical models. *Structural Equation Modeling, 4*, 142–156.

Cheong, J. (2011). Accuracy of estimates and statistical power for testing mediation in latent growth modeling. *Structural Equation Modeling, 18*, 195–211.

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

Cheung, M. W.-L. (2007). Comparison of approaches to constructing confidence intervals for mediating effects using structural equation models. *Structural Equation Modeling, 14*, 227–246.

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

Coffman, D. L., & Zhong, W. (2012). Assessing mediation using marginal structural models in the presence of confounding and moderation. *Psychological Methods, 17*, 642–664.

Cole, S. R., & Hernán, M. A. (2008). Constructing inverse probability weights for marginal structural models. *American Journal of Epidemiology, 168*, 656–664.

Cole, D. A., & Preacher, K. J. (2014). Manifest variable path analysis: Potentially serious and misleading consequences due to uncorrected measurement error. *Psychological Methods, 19*, 300–315.

Cox, M. G., Kisbu-Sakarya, Y., Miočević, M., & MacKinnon, D. P. (2013). Sensitivity plots for confounder bias in the single mediator model. *Evaluation Review, 37*, 405–431.

Craig, C. C. (1936). On the frequency function of *xy*. *Annals of Mathematical Statistics, 7*, 1–15.

Daniel, R. M., De Stavola, B. L., Cousens, S. N., & Vansteelandt, S. (2015). Causal mediation analysis with multiple mediators. *Biometrics, 71*, 1–14.

De Stavola, B. L., Daniel, R. M., Ploubidis, G. B., & Micali, N. (2015). Mediation analysis with intermediate confounding: Structural equation modeling viewed through the causal inference lens. *American Journal of Epidemiology, 181*, 64–80.

Didelez, V., Dawid, A. P., & Geneletti, S. (2006). Direct and

indirect effects of sequential treatments. In *Proceedings of the 22nd Conference on Uncertainty in Artificial Intelligence*, 138–164.

Ding, P., & VanderWeele, T. J. (2016). Sensitivity analysis without assumptions. *Epidemiology, 27*, 368–377.

Duncan, O. D. (1966). Path analysis: Sociological examples. *American Journal of Sociology, 72*, 1–16.

Duncan, O. D., Featherman, D. L., & Duncan, B. (1972). *Socioeconomic background and achievement*. New York: Seminar Press.

Edgington, E. S. (1995). *Randomization tests* (3rd ed.). New York: Marcel-Dekker.

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

Enders, C. K., Fairchild, A. J., & MacKinnon, D. P. (2013). A Bayesian approach to estimating mediation effects with missing data. *Multivariate Behavioral Research, 48*, 340–369.

Falk, C. F., & Biesanz, J. C. (2015). Inference and interval estimation methods for indirect effects with latent variable models. *Structural Equation Modeling, 22*, 24–38.

Folmer, H. (1981). Measurement of the effects of regional policy instruments by means of linear structural equation models and panel data. *Environment and Planning A, 13*, 1435–1448.

Fox, J. (1980). Effect analysis in structural equation models: Extensions and simplified methods of computation. *Sociological Methods and Research, 9*, 3–28.

Frangakis, C. E., & Rubin, D. B. (2002). Principal stratification in causal inference. *Biometrics, 58*, 21–29.

Fritz, M. S., Kenny, D. A., & MacKinnon, D. P. (2016). The combined effects of measurement error and omitting confounders in the single-mediator model. *Multivariate Behavioral Research, 51*, 681–697.

Georgeson, A. R., Valente, M. J., & Gonzalez, O. (2021). Evaluating response shift in statistical mediation analysis. *Advances in Methods and Practices in Psychological Science, 4*, 1–13.

Goetgeluk, S., Vansteelandt, S., & Goetghebeur, E. (2008). Estimation of controlled direct effects. *Journal of the Royal Statistical Society B: Statistical Methodology, 70*, 1049–1066.

Goldberg, L., MacKinnon, D. P., Elliot, D., Moe, E., Clarke, G., & Cheong, J. (2000). Adolescents Training and Learning to Avoid Steroids program: Preventing drug use and promoting health behaviors among adolescent athletes: Results of the program. *Archives of Pediatric and Adolescent Medicine, 154*, 332–338.

Gonzalez, O., & MacKinnon, D. P. (2018). A bifactor approach to model multifaceted constructs in statistical mediation analysis. *Educational and Psychological Measurement, 78*, 5–31.

Gonzalez, O., & MacKinnon, D. P. (2021). The measurement of the mediator and its influence on statistical mediation conclusions. *Psychological Methods, 26*, 1–17.

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

Graff, J., & Schmidt, P. (1982). A general model for decomposition of effects. In K. G. Jöreskog & H. Wold (Eds.) *Systems under indirect observation: Causality, structure, prediction* (pp. 131–148). Amsterdam: North-Holland.

Hernán, M. A., & Robins, J. M. (2006). Estimating causal effects from epidemiological data. *Journal of Epidemiology and Community Health, 60*, 578–586.

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

Holland, P. W. (1988b). Comment: Causal mechanism or causal effect: Which is best for statistical science? *Statistical Science, 3*, 186–188.

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

Imai, K., Keele, L., & Tingley, D. (2010). A general approach to causal mediation analysis. *Psychological Methods, 15*, 309–334.

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

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

Imbens, G. W., & Rubin, D. B. (2015). *Causal inference in statistics, social, and biomedical sciences.* Cambridge, UK: Cambridge University Press.

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

Joffe, M. M., & Greene, T. (2009). Related causal frameworks for surrogate outcomes. *Biometrics, 65*, 530–538.

Jöreskog, K. G. (1970). A general method for analysis of covariance structures. *Biometrika, 57*, 239–251.

Jöreskog, K. G. (1973). A general method for estimating a linear structural equation system. In A. S. Golberger & O. D. Duncan (Eds.), *Structural equation models in the social sciences* (pp. 85–112). New York: Seminar Press.

Jöreskog, K. G., & Sörbom, D. (2006). *LISREL* [Computer program]. Chicago: Scientific Software International, Inc.

Kendall, P. L., & Lazarsfeld, P. F. (1950). Problems of survey analysis. In R. K. Merton & P. F. Lazarsfeld (Eds.), *Continuities in social research: Studies in the scope and method of "The American Soldier"* (pp. 133–196). Glencoe, IL: Free Press.

Kenny, D. A., Kashy, D. A., & Bolger, N. (1998). Data analysis in social psychology. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *The handbook of social psychology* (4th

ed., Vol. 1, pp. 233–265). New York: Oxford University Press.

Kisbu-Sakarya, Y., Mackinnon, D. P., Valente, M. J., & Çetinkaya, E. (2020). Causal mediation analysis in the presence of post-treatment confounding variables: A Monte Carlo simulation study. *Frontiers in Psychology, 11*, Article 2067.

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

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

Last, J. M. (1988). *A dictionary of epidemiology* (2nd ed.). New York: Oxford University Press.

Lazarsfeld, P. F. (1955). Interpretation of statistical relations as a research operation. In P. F. Lazarsfeld & M. Rosenberg (Eds.), *The language of social research: A reader in the methodology of social research* (pp. 115–125). Glencoe, IL: Free Press.

Ledgerwood, A., & Shrout, P. E. (2011). The trade-off between accuracy and precision in latent variable models of mediation processes. *Journal of Personality and Social Psychology, 101*, 1174–1188.

Liu, X., & Wang, L. (2021). The impact of measurement error and omitting confounders on statistical inference of mediation effects and tools for sensitivity analysis. *Psychological Methods. 6*, 327–342.

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

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

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

MacKinnon, D. P., & Lamp, S. J. (2021). A unification of mediation, confounder, and collider effects. *Prevention Science, 22*, 1185–1193.

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., Valente, M. J., & Gonzalez, O. (2020). The correspondence between causal and traditional mediation analysis: The link is the mediator by treatment interaction. *Prevention Science, 21*, 147–157.

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

Mauro, R. (1990). Understanding LOVE (left out variables error): A method for estimating the effects of omitted variables. *Psychological Bulletin, 108*, 314–329.

Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13– 103). New York: American Council on Education/Macmillan.

Meeker, W. Q., Jr., Cornwell, L. W., & Aroian, L. A. (1981). The product of two normally distributed random variables. In W. J. Kennedy & R. E. Odeh (Eds.), *Selected tables in mathematical statistics, Vol. VII* (pp. 1–256). Providence, RI: American Mathematical Society.

Meeker, W. Q., & Escobar, L. A. (1995). Teaching about approximate confidence regions based on maximum likelihood estimation. *American Statistician, 49*, 48–53.

Millsap, R. E. (2011). *Statistical approaches to measurement invariance*. New York: Routledge.

Miočević, M., MacKinnon, D. P., & Levy, R. (2017). Power in Bayesian mediation analysis for small sample research. *Structural Equation Modeling, 24*, 666–683.

Moerkerke, B., Loeys, T., & Vansteelandt, S. (2015). Structural equation modeling versus marginal structural modeling for assessing mediation in the presence of posttreatment confounding. *Psychological Methods, 20*, 204–220.

Mood, A. M., Graybill, F. A., & Boes, D. C. (1974). *Introduction to the theory of statistics* (3rd ed.). New York: McGraw-Hill.

Mosteller, F., & Tukey, J. W. (1977). *Data analysis and regression: A second course in statistics*. Reading, MA: Addison-Wesley.

Muthén, L. K., & Muthén, B. O. (1998/2014). *Mplus User's guide*. Los Angeles: Authors.

Muthén, B. O., Muthén, L. K., & Asparouhov, T. (2017). *Regression and mediation analysis using Mplus*. Los Angeles, CA: Muthén & Muthén.

Naimi, A. I., Moodie, E. E. M., Auger, N., & Kaufman, J. S. (2014). Constructing inverse probability weights for continuous exposures. *Epidemiology, 25*, 292–299.

Neale, M. C., & Miller, M. B. (1997). The use of likelihood-based confidence intervals in genetic models. *Behavior Genetics, 27*, 113–120.

Oort, F. J., Visser, M. R., & Sprangers, M. A. (2009). Formal definitions of measurement bias and explanation bias clarify measurement and conceptual perspectives on response shift. *Journal of Clinical Epidemiology, 62*, 1126–1137.

Olivera-Aguilar, M., Rikoon, S., Gonzalez, O., Kisbu-Sakarya, Y., & MacKinnon, D. P. (2018). Bias, Type 1 error rates, and statistical power of a latent mediation model in the presence of violations of invariance. *Educational and Psychological Measurement, 78*, 460–481.

Pearl, J. (2001). Direct and indirect effects. In J. Breese & D. Koller (Eds.), *Proceedings of the 17th Conference on Uncertainty in Artificial Intelligence* (pp. 411–420). San Francisco: Morgan Kaufmann.

Pearl, J. (2009). *Causality* (2nd ed.). New York: Cambridge University Press.

Pearl, J. (2012). The causal mediation formula—a guide to the assessment of pathways and mechanisms. *Prevention Science, 13*, 426–436.

Pearl, J. (2014). Interpretation and identification of causal mediation. *Psychological Methods, 19*, 459–481.

Pek, J., & Hoyle, R. H. (2016). On the (in)validity of tests of simple mediation: Threats and solutions. *Social and Personality Psychology Compass, 10*, 150–163.

Petersen, M. L., Sinisi, S. E., & van der Laan, M. J. (2006). Estimation of direct causal effects. *Epidemiology, 17*, 276–284.

Preacher, K. J., & Hayes, A. F. (2004). SPSS and SAS procedures for estimating indirect effects in simple mediation models. *Behavior Research Methods, Instruments, and Computers, 36*, 717–731.

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

Preacher, K. J., Zyphur, M. J., & Zhang, Z. (2010). A general multilevel SEM framework for assessing multilevel mediation. *Psychological Methods, 15*, 209–233.

Rijnhart, J. J. M., Valente, M. J., MacKinnon, D. P., Twisk, J. W. R., & Heymans, M. W. (2021). The use of traditional and causal estimators for mediation models with a binary outcome and exposure–mediator interaction. *Structural Equation Modeling, 28*, 345–355.

Robins, J. (1986). A new approach to causal inference in mortality studies with a sustained exposure period—application to control of the healthy worker survivor effect. *Mathematical Modelling, 7*, 1393–1512.

Robins, J. M. (2000). Marginal structural models versus structural nested models as tools for causal inference. In M. E. Halloran & D. Berry (Eds.), *Statistical models in epidemiology: the environment and clinical trials* (pp. 95–134). New York: Springer-Verlag.

Robins, J. M., & Greenland, S. (1992). Identifiability and exchangeability for direct and indirect effects. *Epidemiology, 3*, 143–155.

Robins, J. M., Hernán, M. A., & Brumback, B. (2000). Marginal structural models and causal inference in epidemiology. *Epidemiology, 11*, 550–560.

Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika, 7*, 41–55.

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software, 48*, 1–36.

Smith, L. H., & Vanderweele, T. J. (2019). Mediational E-values: Approximate sensitivity analysis for unmeasured mediator-outcome confounding. *Epidemiology, 30*, 835–837.

Sobel, M. E. (1982). Asymptotic confidence intervals for indirect effects in structural equation models. *Sociological Methodology, 13*, 290–312.

Sobel, M. E. (1986). Some new results on indirect effects and their standard errors in covariance structure models. *Sociological Methodology, 16*, 159–186.

Sobel, M. E. (1987). Direct and indirect effects in linear structural equation models. *Sociological Methods and Research, 16*, 155–176.

Sprangers, M. A., & Schwartz, C. E. (1999). Integrating response shift into health-related quality of life research: a theoretical model. *Social Science and Medicine, 48*, 1507–1515.

Springer, M. D., & Thompson, W. E. (1966). The distribution of products of independent random variables. *SIAM Journal on Applied Mathematics, 14*, 511–526.

Steen, J., Loeys, T., Moerkerke, B., & Vansteelandt, S. (2017). Flexible mediation analysis with multiple mediators. *American Journal of Epidemiology, 2*, 95–115.

Stone, C. A., & Sobel, M. E. (1990). The robustness of estimates of total indirect effects in covariance structure models estimated by maximum likelihood. *Psychometrika, 55*, 337–352.

Taylor, A. B., MacKinnon, D. P., & Tein, J. Y. (2008). Tests of the three-path mediated effect. *Organizational Research Methods, 11*, 241–269.

Tchetgen, E. J. T., & VanderWeele, T. J. (2014). On identification of natural direct effects when a confounder of the mediator is directly affected by exposure. *Epidemiology, 25*, 282–291.

Ten Have, T. R., Joffe, M. M., Lynch, K. G., Brown, G. K., Maisto, S. A., & Beck, A. T. (2007). Causal mediation analyses with rank preserving models. *Biometrics, 63*, 926–934.

Tingley, D., Yamamoto, T., Hirose, K., Keele, L., & Imai, K. (2014). Mediation: R Package for causal mediation analysis. *Journal of Statistical Software, 59*, 1–38.

Tofighi, D., & MacKinnon, D. P. (2011). RMediation: An R package for mediation analysis confidence intervals. *Behavior Research Methods, 43*, 692–700.

Tofighi, D., & MacKinnon, D. P. (2016). Monte Carlo confidence intervals for complex functions of indirect effects. *Structural Equation Modeling, 23*, 194–205.

Valente, M. J., Gonzalez, O., Miočević, M., MacKinnon, D. P. (2016). A note on testing mediated effects in structural equation models: Reconciling past and current research on the performance of the test of joint significance. *Educational and Psychological Measurement, 76*, 889–911.

Valente, M. J., MacKinnon, D. P., & Mazza, G. L. (2020). A viable alternative when propensity scores fail: Evaluation of inverse propensity weighting and sequential G-estimation in a two-wave mediation model. *Multivariate Behavioral Research, 55*, 165–187.

Valente, M. J., Rijnhart, J. J. M., Smyth, H. L., Muniz, F. B., & MacKinnon, D. P. (2020). Causal mediation programs in R, Mplus, SAS, SPSS, and Stata. *Structural Equation Modeling, 27*, 975–984.

Valeri, L., & VanderWeele, T. J. (2013). Mediation analysis allowing for exposure–mediator interactions and causal

interpretation: Theoretical assumptions and implementation with SAS and SPSS macros. *Psychological Methods, 18*, 137–150.

Valeri, L., & VanderWeele, T. J. (2015). SAS macro for causal mediation analysis with survival data. *Epidemiology, 26*, e23–e24.

VanderWeele, T. J. (2008). Simple relations between principal stratification and direct and indirect effects. *Statistics and Probability Letters, 78*, 2957–2962.

VanderWeele, T. J. (2009). Marginal structural models for the estimation of direct and indirect effects. *Epidemiology, 20*, 18–26.

VanderWeele, T. J. (2010). Bias formulas for sensitivity analysis for direct and indirect effects. *Epidemiology, 21*, 540–551.

VanderWeele, T. J. (2014). A unification of mediation and interaction: A 4-way decomposition. *Epidemiology, 25*, 749–761.

VanderWeele, T. J., & Robins, J. M. (2008). Empirical and counterfactual conditions for sufficient cause interactions. *Biometrika, 95*, 49–61.

VanderWeele, T. J., & Vansteelandt, S. (2009). Conceptual issues concerning mediation, interventions and composition. *Statistics and Its Interface, 2*, 457–468.

Vanderweele, T. J., & Vansteelandt, S. (2010). Odds ratios for mediation analysis for a dichotomous outcome. *American Journal of Epidemiology, 172*, 1339–1348.

Vanderweele, T. J., Vansteelandt, S., & Robins, J. M. (2014). Effect decomposition in the presence of an exposure-induced mediator–outcome confounder. *Changes, 29*, 997–1003.

Vansteelandt, S. (2009). Estimating direct effects in cohort and case–control studies. *Epidemiology, 20*, 851–860.

Wiley, D. E. (1973). The identification problem for structural equation models with unmeasured variables. In A. S. Goldberger & O. D. Duncan (Eds.), *Structural equation models in the social sciences* (pp. 69–83). New York: Seminar Press.

Wolfle, L. M. (1999). Sewall Wright on the method of path coefficients: An annotated bibliography. *Structural Equation Modeling, 6*, 280–291.

Wright, S. (1920). The relative importance of heredity and environment in determining the piebald pattern of guinea-pigs. *Proceedings of the National Academy of Sciences USA, 6*, 320–332.

Wright, S. (1921). Correlation and causation. *Journal of Agricultural Research, 20*, 557–585.

Wright, S. (1923). The theory of path coefficients: A reply to Niles's criticism. *Genetics, 8*, 239–255.

Yuan, Y., & MacKinnon, D. P. (2009). Bayesian mediation analysis. *Psychological Methods, 14*, 301–322.