Advances in Mediation Analysis: A Survey and Synthesis of New Developments

Kristopher J. Preacher

Department of Psychology and Human Development, Vanderbilt University, Nashville, Tennessee 37203-5721; email: kris.preacher@vanderbilt.edu

Annu. Rev. Psychol. 2015. 66:825–52

First published online as a Review in Advance on August 21, 2014

The *Annual Review of Psychology* is online at psych.annualreviews.org

This article's doi:
10.1146/annurev-psych-010814-015258

Keywords

indirect effect, longitudinal mediation, causal inference, discrete variables, multilevel modeling

Abstract

Mediation processes are fundamental to many classic and emerging theoretical paradigms within psychology. Innovative methods continue to be developed to address the diverse needs of researchers studying such indirect effects. This review provides a survey and synthesis of four areas of active methodological research: (*a*) mediation analysis for longitudinal data, (*b*) causal inference for indirect effects, (*c*) mediation analysis for discrete and nonnormal variables, and (*d*) mediation assessment in multilevel designs. The aim of this review is to aid in the dissemination of developments in these four areas and suggest directions for future research.

Contents

INTRODUCTION

Mediation analysis is a family of methods designed to extract information about the causal mechanism(s) by which a predictor affects an outcome. Mediation is at the heart of many classic and emerging theoretical paradigms within psychology. For example, Narayan et al. (2013) found that externalizing behavior mediated the effect of exposure to interparental violence on dating violence. Newland et al. (2013) found that both depression and anxiety mediated the effect of economic pressure on maternal sensitive parenting. Mediation analysis is popular throughout the social sciences—and especially in prevention and medical research, where it is of interest to determine the mechanism(s) by which a treatment exerts its effect.

Many overviews of mediation analysis exist (e.g., Hayes 2013; MacKinnon 2008; MacKinnon & Fairchild 2009; MacKinnon et al. 2002, 2007, 2013a,b). In their review of the state of the art in mediation analysis, MacKinnon et al. (2007) focused primarily on the three-variable mediation model, concentrating on questions of estimation and statistical inference. Several other topics and extensions were also addressed, including effect size, multilevel mediation, mediation with categorical outcomes, multiple mediator models, longitudinal mediation, combining mediation with moderation, and causal inference. MacKinnon (2008) expanded on these topics and more. In the intervening years, questions of statistical inference in mediation analysis have, by and large, been answered. For example, several well-performing methods have been suggested for constructing confidence intervals (CIs) for indirect effects, including several bootstrap CI methods

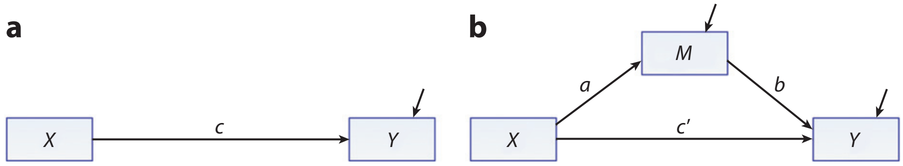

Figure 1

(*a*) Regression of Y on X. (*b*) The simple mediation model with M as a mediator of the effect of X on Y.

(Bollen & Stine 1990, Shrout & Bolger 2002), Bayesian credible intervals (Yuan & MacKinnon 2009), Monte Carlo CIs (MacKinnon et al. 2004, Preacher & Selig 2012), and methods involving product distributions (MacKinnon et al. 2004). These methods agree more often than they disagree (Hayes & Scharkow 2013); thus, which method is chosen is often of little consequence. Several software tools are now widely available to help researchers implement each of these methods, bringing sophisticated forms of mediation analysis within the reach of any social scientist.

In the brief period since the publication of MacKinnon et al. (2007), the focus of the mediation literature has largely shifted away from questions of statistical inference and toward exploring ways to improve the match between researchers' theoretical questions and the statistical methods used to investigate them. In particular, four of the extensions to the single mediation model explored by MacKinnon et al. (2007) have come to dominate the mediation literature. These are singled out here for greater attention: (*a*) mediation analysis for longitudinal data, (*b*) causal inference for indirect effects, (*c*) mediation analysis for discrete and nonnormal variables, and (*d*) mediation assessment in multilevel designs. The goal of this review is to summarize and synthesize key points and developments in the diverse literature on these topics. First, simple mediation analysis is reviewed. In subsequent sections, the four extensions are addressed.

SIMPLE MEDIATION

Consider the simple regression model in **Figure 1*a***. X is the predictor (independent variable) and Y is the outcome (dependent variable). The regression weight c represents the total effect of X on Y.

$$Y_i = i_1 + cX_i + e_{1i} \tag{1}$$

Most treatments of mediation analysis begin with simple mediation, a three-variable causal system in **Figure 1*b***. M is the mediator (or intervening variable) hypothesized to mediate the effect of X on Y (c).

$$\begin{aligned} Y_i &= i_2 + c'X_i + bM_i + e_{2i} \\ M_i &= i_3 + aX_i + e_{3i}, \end{aligned} \tag{2}$$

and c' is the direct effect of X on Y after controlling for M. Evidence that M serves as a mediator is often said to exist when the indirect effect (the product of slopes a and b) is statistically and practically significant. Covariates may also be included, and although the effects are usually specified as linear in practice, they can be nonlinear (Hayes & Preacher 2010, Lockhart 2012, Stolzenberg 1980).

MEDIATION MODELS FOR LONGITUDINAL DATA

The traditional trivariate mediation model in **Figure 1** has held sway for decades in studies applying mediation analysis. However, if the goal is to draw inferences about a causal process, it is not sufficient to merely show that variables are characterized by a theoretically compelling pattern of relationships. At a minimum, some time should elapse between a putative cause and its associated effect to allow for the effect to occur or unfold. In response to psychologists' pervasive reliance on cross-sectional designs and models for assessing mediation, methodologists have developed models that respect the role of time, but in different ways. Three major classes of longitudinal mediation models are in common use: methods based on the cross-lagged panel model (CLPM), the latent growth curve model (LGM), and the latent change score (LCS) model (Bentley 2011; MacKinnon 2008; MacKinnon et al. 2007, 2013b; Roth & MacKinnon 2012; Selig & Preacher 2009). These are examined below, followed by a discussion of emerging methodologies that herald the arrival of yet more causally defensible methods for studying mediation. In addition to strengthening causal inference, longitudinal mediation models allow greater latitude in theory testing and grant the ability to investigate the stability of effects over time as well as to build evidence for a particular causal ordering of variables and to study whether change itself plays a role in a mediation process.

CLPM: cross-lagged panel model

LGM: latent growth curve model

LCS: latent change score

SEM: structural equation modeling

Panel Models

It seems reasonable that claims of causality should be strengthened by deliberately staggering the assessment of *X*, *M*, and *Y* (e.g., the process $X_{t-2} \rightarrow M_{t-1} \rightarrow Y_t$, or sequential design; Mitchell & Maxwell 2013). However, while X_{t-2} is affecting M_{t-1} and M_{t-1} is affecting Y_t, M_{t-2} also may affect M_{t-1} and Y_{t-1} also may affect Y_t. Inclusion of such autoregressive effects is important because they reflect the stability of individual differences in a variable over the chosen lag (time elapsed between assessments). Presumably, only the unstable variance in a variable at time *t* may be explained by other variables assessed at earlier occasions, so it is sensible to include prior assessments of a variable to partial out this stable variance. Moreover, the size of effects (autoregressive or otherwise) usually will vary as a function of lag (Cole & Maxwell 2003; Gollob & Reichardt 1987, 1991; Maxwell & Cole 2007; Maxwell et al. 2011). Typically, there is no one "correct" lag at which to assess an effect; rather, effects tend to vary as a function of lag, and a more complete understanding of relationships among variables can result from assessing them at multiple lags (Selig et al. 2012).

A popular choice for assessing longitudinal mediation is the CLPM, which is based on structural equation modeling (SEM) for repeated measures of *X*, *M*, and *Y* in which each variable depends not only on causally prior variables but also on prior assessments of the same variable (Gollob & Reichardt 1991). The parameters of a CLPM most relevant for mediation analysis are those that connect different variables across measurement occasions separated by the chosen lag. There are several variations on this model. An example CLPM is depicted in **Figure 2*a***, in which parameters with similar interpretations across time are constrained to equality. The indirect effect of X_{t-2} on Y_t via M_{t-1} is $a \times b$. More than three waves of measurement may be included. Individuals are assumed to be assessed at roughly the same measurement occasions, but this assumption can be relaxed if lag is explicitly included as a moderator variable (Selig et al. 2012).

The interpretation and generalizability of indirect effects identified using CLPM depend on the degree to which stability, stationarity, and equilibrium hold (Cole & Maxwell 2003, MacKinnon 2008). Stability refers to the degree to which individual differences in a variable are maintained over time. Stationarity is descriptive of a causal structure that remains unchanged over time. Finally, equilibrium implies that cross-sectional variances and covariances remain stable over time. The biases that can result from using cross-sectional and sequential designs are further explored by Maxwell & Cole (2007) and Mitchell & Maxwell (2013).

Specifying a CLPM requires at least three occasions of measurement. However, an indirect effect linking three variables may be estimated with data collected at only two occasions by using the half-longitudinal design (**Figure 2*b***; Cole & Maxwell 2003). The effect of X_{t-1} on M_t is multiplied by the effect of M_{t-1} on Y_t to yield the indirect effect. Finally, for situations in

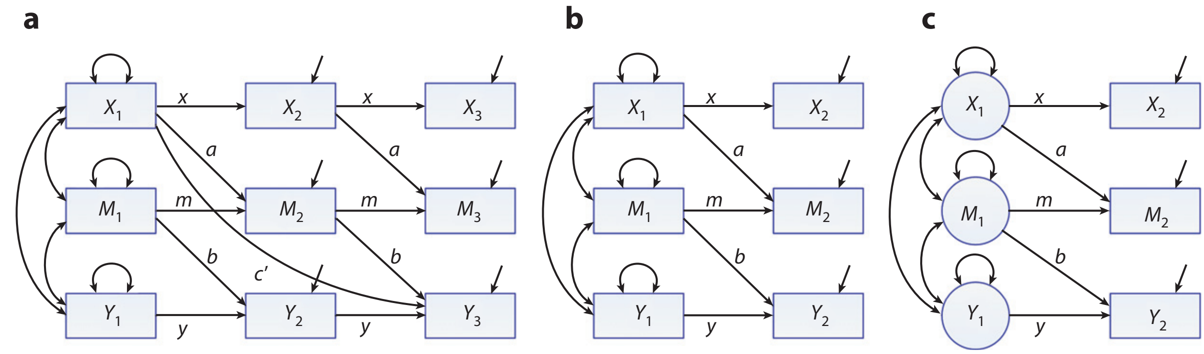

Figure 2

(*a*) Full cross-lagged panel model. (*b*) Cross-lagged panel model for a half-longitudinal design. (*c*) Latent longitudinal model. Although not depicted, residuals are typically permitted to covary across variables within a given occasion. Paths *a*, *b*, and *c′* are cross-lag paths. Paths *x*, *m*, and *y* are autoregressive effects.

which only cross-sectional data are available, Gollob & Reichardt (1987, 1991) describe a latent longitudinal model (**Figure 2*c***) that could be used to investigate mediation if the researcher is willing to make several additional assumptions based on prior information, imposed in the form of identifying constraints. Having fewer repeated measures requires heavier reliance on untestable assumptions. For example, moving from the CLPM to the half-longitudinal design sacrifices the ability to test stationarity. Moving to the latent longitudinal design further sacrifices the ability to assess stability and test equilibrium.

Latent Growth Curve Models

The CLPM addressed the questions of whether, and to what degree, individual differences in *X* predict variability in *Y* via *M*, with advantages accruing from the use of repeated measures of the same variables on the same individuals. An alternative method has been suggested in the context of the LGM (Bollen & Curran 2006, Cheong 2011, Cheong et al. 2003). The LGM permits aspects of longitudinal change in a variable (e.g., individuals' intercepts and slopes) to assume the role of *X*, *M*, or *Y* in a mediation model. For example, it might be of interest to determine the degree to which a teaching intervention (*X*) influences linear change in math ability in high school (*Y*) via the rate of skill acquisition in elementary school (*M*). In this example, both the mediator and outcome are individuals' rates of change, or linear slopes over time, whereas *X* is a binary predictor. Variations on this theme may be imagined, such as models in which participants are randomized to an intervention (*X*), both the intercept and slope of a repeatedly measured variable serve as joint mediators, and *Y* is a distal outcome (see **Figure 3**).

Important caveats are associated with use of LGM mediation models. First, if two variables (say, *M* and *Y*) are assessed over the same span of time, their growth factors will not adhere to a clear temporal order, and causal priority is murky at best (MacKinnon 2008, MacKinnon et al. 2007, Selig & Preacher 2009, von Soest & Hagtvet 2011). This problem may be ameliorated somewhat by using a two-stage piecewise LGM such that early change in *M* will temporally precede later change in *Y* (Bentley 2011). Second, the role of time in the LGM mediation model is quite different from the role of time in the CLPM mediation model. In the CLPM, time serves to determine the lag separating repeated measures, and effects can vary as a function of lag. In the LGM mediation model, the lag between repeated measures of the same variable is not as important. But the lag

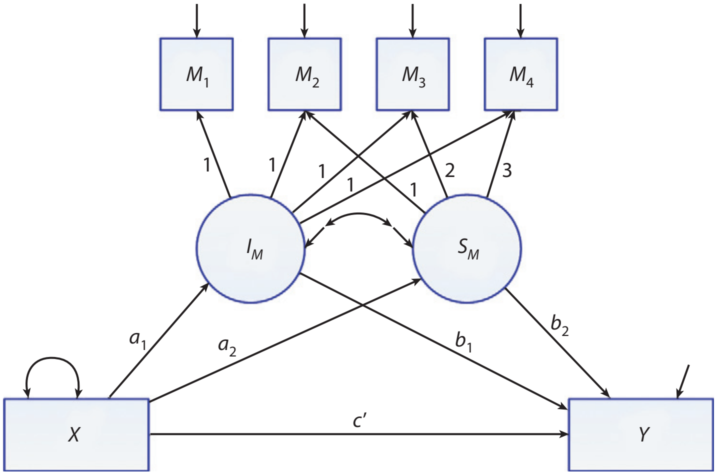

Figure 3

One example of a latent growth curve mediation model. $M_1 - M_4$ are repeated measures of a mediator, and I_M and S_M are the intercept and slope factors for linear growth in $M_1 - M_4$. In this model, I_M and S_M act as joint mediators of the effect of X on Y.

separating repeated measures of different variables (say, M versus Y) will influence the structural paths important for mediation. Third, care must be taken when interpreting indirect effects in such models when X and M are aspects of change; because intercepts and slopes typically covary, their effects on other variables are mutually conditioned on the other's inclusion in the model (von Soest & Hagtvet 2011). Moreover, the intercept-slope covariance depends on the often arbitrary decision of where to center the time variable.

One recent advancement in the application of growth modeling to mediation involves nonlinear functions (e.g., exponential trends), with aspects of change acting as variables involved in a mediation model (Fritz 2014). For example, X may be a time-invariant treatment assignment, and M and Y each may change exponentially over time toward an asymptote. It is possible to assess the effect of X on aspects of change in M (e.g., the instantaneous rate of change) while M is treated as a time-varying lagged predictor of Y.

Latent Change Score Models

MacKinnon (2008) suggests an adaptation of the LCS model (also called the latent difference score model) for studying longitudinal mediation. The basic idea behind LCS models is to use latent variables to represent the difference (change) between adjacent measurements of a repeatedly measured variable (McArdle 2001). This latent difference is interpretable as the rate of change between occasions $t-1$ and t. Typically, an LCS model is fit to several sequential pairs of measurements. In the mediation context, the latent difference variables may participate as variables in a mediation model. Unlike the CLPM, LCS models do not address the relationships among the variables themselves over time but rather focus on change in a variable and the relationships among these changes. Unlike the mediation LGM, change is not assumed to be constant across lags. Furthermore, unlike the mediation LGM, change occurs across only a single measurement period (although more elaborate LCS models are possible).

Emerging Modeling Strategies for Longitudinal Mediation

There has recently been an influx of several new ideas about how to incorporate time into mediation analysis in ways that are arguably more consistent with theory and with our conceptual understanding of how mediation processes work. Two of these novel methods are described in this section.

SSM: state space modeling

CTM: continuous time model

State space models. State space modeling (SSM) may be used to assess mediation with longitudinal data using either a single person's multivariate time series (many repeated measures of X, M, and Y for one person) or multiple persons' time series (Gu et al. 2014). SSM involves a measurement equation, linking observations of X, M, and Y at occasion t to latent variables at occasion t, and a transition equation, depicting temporal relationships among those latent variables over time. Because parameter estimation is based on as little as one person's data, SSM requires a much larger number of repeated measures than do other methods. However, unlike longitudinal mediation models reviewed above, this model more explicitly recognizes that mediation is a within-person process and allows investigation of whether the indirect effect differs across persons.

Continuous time models. The CLPM model discussed previously limits generalizability to the specific lags chosen to separate the repeated assessments of X, M, and Y. An alternative continuous time model (CTM), suggested by MacKinnon et al. (2007) and Maxwell & Cole (2007) and implemented by P.R. Deboeck and K.J. Preacher (unpublished manuscript) and Deboeck et al. (2013), represents the data-generating process using parameters that are independent of lag. The CLPM is conceptually a special case of the CTM because it is possible to derive the values of the parameters of a CLPM at any given lag in the CTM, not just the lag at which the data were collected. With this capability, it is possible to plot effects linking X, M, and Y as a function of lag to more fully understand the data-generating process. This, in turn, allows estimating the lag associated with the maximum indirect effect. Moreover, CTMs require the same kind of discrete-time data that are used with CLPM.

Discussion

Currently, CLPM is the most often used mediation model for longitudinal data, followed by LGM and then LCS. This may reflect that these models are the most established, and all three can be fit within the standard SEM framework using popular SEM software. The SSM and CTM require specialized software and more specialized knowledge (e.g., experience with time series modeling and CTMs).

The three more established models continue to be improved upon. For example, the CLPM and LGM mediation models have been extended in a number of ways (e.g., CLPM models in which there exist reciprocal effects linking M and Y over time). Simultaneously, the growing interest in the SSM and CTM models reflects an increasing desire in the methodological literature on longitudinal mediation to treat time more dynamically. For example, rather than base a model on between-person individual differences, SSM explicitly models mediation as a process that unfolds over time within an individual. The CTM explicitly acknowledges that effects relating variables in a mediation model usually are not tied to discrete events but rather occur on a continuous basis, with observable effects varying in strength with the lag separating the discretely timed observations of this process.

The SSM and CTM methods are not direct extensions of more established longitudinal mediation models; rather, they return to the conceptual notion of mediation and adopt different assumptions (e.g., substituting assumptions about continuous time for assumptions about discrete time, and within-person change for between-person change). Users of longitudinal mediation models are tasked with philosophically assessing how their theoretical goals may be met by these differing modeling frameworks. Additionally, more work is needed to adapt the CLPM, LCS, SSM, and CTM methods for longitudinal mediation to handle individually varying occasions of measurement; the LGM does not require time-balanced data.

CAUSAL INFERENCE FOR INDIRECT EFFECTS[1]

It is now widely understood that a significant indirect effect is not particularly meaningful if it is based on cross-sectional data. The longitudinal mediation methods previously described emphasize that mediation processes require time to unfold, and they explicitly incorporate time to strengthen causal inference. However, inferring that an effect is causal requires more than establishing the correct temporal ordering of one's variables (Valeri & VanderWeele 2013). Basic requirements for establishing causality include not only (*a*) temporal precedence (the hypothesized cause must precede the effect in time) but also (*b*) observed covariation (the hypothesized cause should be statistically associated with the effect) and (*c*) elimination of plausible alternative explanations for an observed relationship (Shadish et al. 2002). Hence, the literature on causal inference implicitly presupposes longitudinal data (although the intricacies of longitudinal data and designs are rarely addressed).

The causal mediation literature has developed to formalize the requirements for causal inference in tests of mediation hypotheses using both design-based and model-based traditions, described below. A distinct yet compatible perspective is offered by Judea Pearl (see Pearl 2009, 2010; Shadish 2010; Shadish & Sullivan 2012; West & Thoemmes 2010). The design-based tradition seeks to strengthen causal inference by employing experimental design principles, whereas the model-based tradition emphasizes defining effects and identification of assumptions necessary for causal inference. Next, the main tenets of the design-based and model-based traditions are reviewed, along with their prominent implementation strategies and limitations and some new developments.

Design-Based Tradition

The design-based tradition, beginning with the influential work of Donald Campbell (e.g., Campbell & Stanley 1963), has long emphasized the importance of strengthening internal validity, the validity of inferences about the causal nature of observed effects (Shadish et al. 2002). In this framework, the gold standard is a high-quality experimental design that minimizes plausible rival hypotheses and other threats to internal validity (Campbell & Stanley 1963, Shadish 2010, Shadish et al. 2002, Shadish & Sullivan 2012, Smith 1982, West & Thoemmes 2010). Placed in the context of mediation analysis, classic designs can be ranked in decreasing order of evidence for causality: (*a*) randomization of both X and M (double-randomization or the experimental-causal-chain strategy; Smith 1982, Spencer et al. 2005), (*b*) quasi-experimental manipulation of both X and M, (*c*) randomization of X only, (*d*) quasi-experimental manipulation of X only, and (*e*) nonexperimental studies, in which X, M, and Y are measured (Stone-Romero & Rosopa 2008). Double-randomization is the most direct way to rule out confounds in the $X \rightarrow M$ and $X \rightarrow Y$ and

[1]A more extensive list of references on causal inference in mediation analysis is available as an appendix at the author's website, **http://quantpsy.org**. Only selected sources from this large literature are cited here.

$M \rightarrow Y$ effects. However, even double-randomization—the best of the options—is not without limitations, as follows.

Difficulties

1. Often it is difficult to manipulate the proposed mediator.
2. Even when manipulation of M is possible, X may moderate the $M \rightarrow Y$ effect, or observed effects may characterize only a subset of the population (Bullock et al. 2010, Spencer et al. 2005).
3. Individual differences in M arising as a function of X and those arising as a result of manipulation may not be alike in degree or kind (Bullock et al. 2010, Kenny 2008, Stone-Romero & Rosopa 2011).
4. Manipulation targeting the mediator must target only that mediator and no other mediator (Bullock et al. 2010).
5. Unless the $X \rightarrow M$ and $M \rightarrow Y$ effects are homogenous across units, it is possible to observe an average indirect effect that is highly biased with respect to the actual indirect effect (Imai et al. 2013).
6. Because many phenomena worth studying are not subject to manipulation for practical or ethical reasons, much of psychological science would have to be abandoned if only experimental evidence were accepted as scientifically sound (Kenny 2008).
7. If participants are randomized to only one level of X, we observe a value of M or Y only under that level of X, and not under other levels of X.

Implementation strategies. Several implementation strategies are available for addressing these difficulties with double-randomization.

1. *Specificity and consistency designs.* Specificity designs address difficulties 1, 4, and 6 by identifying causal effects through some mediators but not others (multiple mediator, blockage, and enhancement designs; MacKinnon 2008; MacKinnon et al. 2012, 2013). Consistency designs address difficulties 2, 5, and 6, indirectly supporting causality through exploring the generalizability of mediation conclusions across contexts (e.g., mediation meta-analysis; Shadish 1996). Blockage designs, in particular, hold great promise for supporting causal inference in mediation settings. These designs involve manipulation of X followed by measurement of M, but they add a manipulation that either blocks or does not block M from having an effect on Y. Evidence for mediation exists if X affects M, if M affects Y, and if X affects Y only in the nonblocked condition (Robins & Greenland 1992, Sigall & Mills 1998).

2. *Within-subject experimental designs.* Within-subject experimental designs, or crossover designs, address difficulty 7. In some cases it is possible to not only manipulate X, but also expose participants to all X conditions in a within-subjects design and observe their responses to M and Y under each X condition. Judd et al. (2001) describe analytic methods for establishing mediation and moderation in such designs. In cases where the risk of carryover effects (e.g., priming or practice effects) are minimal, within-subjects mediation designs can dramatically increase statistical power and strengthen causal inference, given that participants serve as their own controls.

Still, despite these advances, there are limits to what design may accomplish. When experimental manipulation of M (or even X) is not feasible, the next best alternative is statistical control to eliminate threats to internal validity. For example, whereas Judd & Kenny (1981) emphasized the importance of experimentally manipulating X, they also recognized the possibility that omitted variables could bias the $M \rightarrow Y$ effect and urged statistical control of alternative explanations for effects. Statistical control of potential confounds is a simple special case of model-based strategies, discussed next.

Potential outcome: the value of an outcome variable that would be observed for case i had case i been assigned to a given treatment condition. In practice, only one potential outcome can be realized; the other potential outcomes are counterfactual

Fundamental problem of causal inference: the problem that case i's response on an outcome may be observed under only one treatment condition

Stable unit treatment value assumption (SUTVA): the assumption that the treatment assignment for case i is not affected by assignments for other cases and that there is no variation in treatment across cases. Under SUTVA, the response of case i depends only on the treatment to which case i was assigned

Model-Based Tradition

The model-based tradition owes its origins to the pioneering work of Donald Rubin (1974, 2004); its application to the study of mediation is due primarily to Holland (1986, 1988) and Robins & Greenland (1992). In contrast to the design-based tradition, which emphasizes the efficient use of design to maximize internal validity, the focus of the model-based tradition is on formal definition of effects, often in the context of a statistical model, and on identification of the assumptions and conditions under which causality may be inferred.

First, I review Rubin's framework for causal inference in general and subsequently in the context of mediation. Rubin's framework holds that, for a binary X (although the logic extends to continuous X), the effect of X on Y for case i can be defined as the difference between two potential outcomes: one that would be realized if $X_i = 1$ [i.e., $Y_i(1)$] and one that would be realized if $X_i = 0$ [i.e., $Y_i(0)$]. X causes Y if the potential outcomes for Y differ depending on the value of X; that is, if $Y_i(1) \neq Y_i(0)$. However, in ordinary between-subjects designs, only one of $X_i = 1$ or $X_i = 0$ can be realized (termed the fundamental problem of causal inference; Holland 1986, 1988). Thus, the extent to which causality may be inferred depends on the extent to which a set of assumptions about unobserved events is satisfied. Under random assignment, $Y_i(0)$ and $Y_i(1)$ are independent of the condition to which case i is actually assigned (X_i), in which case the mean difference between X groups is:

$$\begin{aligned} &E[Y_i(1)] - E[Y_i(0)] \\ &\quad = E[Y_i(1)|X_i = 1] - E[Y_i(0)|X_i = 0] \\ &\quad = E[Y_i|X_i = 1] - E[Y_i|X_i = 0], \end{aligned} \tag{3}$$

the familiar mean difference in Y between X groups. That is, whereas we cannot determine whether $Y_i(1) \neq Y_i(0)$ for an individual case, it often is possible to determine whether $E[Y_i(1)] \neq E[Y_i(0)]$ on average. This derivation requires the stable unit treatment value assumption (SUTVA), defined below.

In mediation models, there are potential outcomes on M as well ($M_i(0)$ and $M_i(1)$), only one of which can be observed for case i. So observed outcomes are $Y_i(X_i, M_i(X_i))$. All other potential outcomes are counterfactual. Using this framework, the indirect effect for case i may be defined as

$$\delta_i(x) = Y_i(x, M_i(1)) - Y_i(x, M_i(0)) \tag{4}$$

for $x = 0, 1$. That is, $\delta_i(x)$ is the change that would occur in Y when moving from the value of M if $X_i = 0$ to the value of M if $X_i = 1$. $\delta_i(x)$ can never be observed for a given case i, but under SUTVA the average indirect effect is computable as

$$\bar{\delta}(x) = E[Y_i(x, M_i(1)) - Y_i(x, M_i(0))] \tag{5}$$

for $x = 0, 1$. That is, $\bar{\delta}(x)$ is—for a given value of X_i—the average difference between the value of Y_i had we observed the value of M_i obtained under $x = 0$ and the value of Y_i had we observed the value of M_i obtained under $x = 1$. In the regression context and under standard assumptions, this quantity is equivalent to $a \times b$, the traditional indirect effect computed from coefficients in Equation 2.

Assumptions. Core assumptions of the model-based tradition are discussed in many sources (e.g., Coffman & Zhong 2012, Emsley et al. 2010), and the main ones are highlighted here.

Some authors relax some assumptions (e.g., Hafeman & VanderWeele 2011, Vansteelandt & VanderWeele 2012) or add others in specific circumstances. Other assumptions are necessary but implicit (data are longitudinal; Valeri & VanderWeele 2013), or never mentioned because they are virtually always satisfied (VanderWeele & Vansteelandt 2009). Still other general assumptions are commonly invoked for the specific statistical methods used in model fitting (e.g., correct model specification, independent sampling, error homoscedasticity, etc.) but are not discussed here.

Sequential ignorability: The two-part assumption that (*a*) the effect of X on M is not confounded by variables assessed prior to X and (*b*) the effect of M on Y is not confounded by X or by variables assessed prior to X

Propensity score: the estimated probability that case i is assigned to a particular treatment, given a set of observed covariates

1. *Stable unit treatment value assumption.* SUTVA has two parts: no interference between units and no hidden versions of treatments. SUTVA is an extremely important assumption because it limits the proliferation of potential outcomes for case i to (*a*) those conditioned on manipulated variables, (*b*) observations that are themselves conditioned on manipulated variables, and not (*c*) those observed or potentially observed for other cases (Emsley et al. 2010, Rubin 2010).

2. *Additivity.* This is also called the no-interaction assumption. We assume that $\delta_i(x)$ (see Equation 4) does not vary as a function of the condition to which case i is actually assigned (Emsley et al. 2010, Holland 1988, Imai et al. 2010a, Robins & Greenland 1992). Under additivity, the indirect and direct effects sum to the total effect (e.g., c in the linear regression case of Equations 1 and 2) both for individual cases and on average (Pearl 2011; VanderWeele & Vansteelandt 2009, 2010).

3. *Sequential ignorability.* In order to lend $\bar{\delta}(x)$ (see Equation 5) a causal interpretation, we must assume sequential ignorability (Emsley et al. 2010, Imai et al. 2010b), sometimes termed exchangeability or no unmeasured confounding. Sequential ignorability is a two-part assumption that holds if (*a*) the $X \rightarrow M$ and $X \rightarrow Y$ relationships are not confounded by variables assessed prior to X and (*b*) the $M \rightarrow Y$ relationship is not confounded by X or by variables assessed prior to X. If treatment assignment is randomized, X is ignorable and condition (*a*) automatically holds. Imai et al. (2010b) prove that under sequential ignorability, the classic indirect effect (i.e., $a \times b$, as defined in the linear regression case in Equations 1 and 2) is nonparametrically identified (i.e., expressible in terms of estimable parameters; Pearl 2010). Under sequential ignorability, the mediator is effectively (but not literally) randomly assigned, given X and the covariates, and a causal interpretation of indirect effects is justified for any mediation model (Imai et al. 2010a,b).

Difficulties. Part (*b*) of sequential ignorability is a strong assumption that is difficult to test. A common way to address violations of sequential ignorability is to include as covariates potential pretreatment confounders of the $X \rightarrow M$ relationship, the $M \rightarrow Y$ relationship, or both, and to demonstrate that X does not moderate the $M \rightarrow Y$ path (or to model this interaction if it exists).

Because unmeasured confounding (i.e., violation of sequential ignorability) is such a ubiquitous problem (for an overview, see Richiardi et al. 2013), several strategies have been developed from within the model-based framework to cope with it (Emsley et al. 2010). In the following sections, five such methods are described—instrumental variables, propensity scores, marginal structural models, principal stratification, and sensitivity analyses—although other methods also exist (e.g., Lynch et al. 2008, Tchetgen Tchetgen & Lin 2012, Ten Have et al. 2007).

Implementation strategies

1. *Instrumental variables.* One approach to try to satisfy the sequential ignorability assumption is to find an instrumental variable—a variable correlated with M but uncorrelated with the error term of Y (Holland 1988, Raudenbush et al. 2012). Then M is regressed on the instrumental variable, and Y is regressed on the predicted values (two-stage least squares estimation). A limitation is that the instrumental variable method requires the effect of X on Y to be entirely indirect through M (the exclusion restriction). The exclusion restriction usually is not realistic in practice (Imai et al. 2010a, 2011; Jo 2008).

PS: principal stratification

2. *Propensity scores.* A second approach to try to satisfy sequential ignorability is to use propensity scores to reduce or remove the selection bias that results when participants are not randomly assigned to levels of M (Coffman 2011, Jo et al. 2011). The propensity score (π_i) is the probability that an individual receives a particular level of M, given a set of measured confounders Z_i. Propensity scores are commonly obtained via logistic regression M on Z, but there are other methods. The π_i can then be included as covariates in the mediation model. Advantages are that the π_i reduce a large number of measured potential confounders into a single numerical summary, and they can be used to control for posttreatment confounders. Treating the π_i as covariates renders M essentially randomly assigned, assuming all important confounders have been measured.

3. *Marginal structural models.* A third approach to try to satisfy the sequential ignorability assumption is to use marginal structural models (Coffman & Zhong 2012, Lange et al. 2012, VanderWeele 2009). Marginal structural models are models (not necessarily linear regression models) for expected values of potential outcomes rather than for observed outcomes. Specifying models for M and Y in this way requires explicit representation of the important effects in terms of potential outcomes. This method uses inverse probability weights, which in turn are computed using propensity scores. Inverse probability weights are the inverse probabilities of being assigned to the treatment actually received, conditional on potential confounders included in the propensity model. Inverse probability weights are incorporated into the model like survey weights.

4. *Principal stratification.* A fourth approach, principal stratification (PS), satisfies the sequential ignorability assumption when confounding is possible due to nonrandom assignment of M (Frangakis & Rubin 2002, Gallop et al. 2009, Jo 2008, Lynch et al. 2008, Rubin 2004). Considering binary X and M for simplicity, four principal strata may be defined: compliant (concordant) mediators (those for whom $M(1) = 1$ and $M(0) = 0$), always mediators ($M(1) = 1$ and $M(0) = 1$), never mediators ($M(1) = 0$ and $M(0) = 0$), and defiant (discordant) mediators ($M(1) = 0$ and $M(0) = 1$). The direct effect is estimated using only the always mediator and never mediator strata, because for these strata response to treatment assignment does not depend on the mediator. We never fully know whether a given individual is a complier, an always mediator, a never mediator, or a defier. PS models involve estimating the parameters of conditional mixtures for principal stratum membership, aided by including predictors of class membership. PS consists of estimating $X \rightarrow Y$ effects within the principal strata (latent classes) defined by potential outcomes on M. Conceptually, this is like holding M constant while estimating the direct effect of X on Y. Once the direct effect of X is estimated, the indirect effect may be quantified as a function of the parameters of the joint distribution of potential outcomes for M and Y (Elliott et al. 2010, Imai et al. 2011, Jo 2008).

The PS method requires strong predictors of the principal strata. PS does not require the no-interaction assumption (additivity) between X and M, but it does require no interaction between X and baseline covariates within strata. Often, to simplify estimation and inference, it is assumed that defiers do not exist (monotonicity; Gallop et al. 2009, Robins & Greenland 1992) or that the proportion of compliers is less than that of defiers (stochastic monotonicity; Elliott et al. 2010).

5. *Sensitivity analyses.* If sequential ignorability is violated and not all potential confounders were measured and included in the model, a sensitivity analysis can be undertaken to investigate the robustness of results to the influence of omitted confounders of the $M \rightarrow Y$ relationship. For example, because sequential ignorability implies that the residual correlation between M and Y should be 0, the researcher can manipulate this correlation between the extremes of -1 and $+1$ and observe the robustness of the indirect effect results (Imai et al. 2010a,b).

Discussion

Despite many differences, in the specific case of mediation analysis, the model- and design-based traditions for causal inference largely agree on the essentials (Shadish & Sullivan 2012). Both frameworks represent systematic and principled attempts to dissect and understand the cause-effect relationships that underlie mediation. The primary strength of the design-based tradition is its focus on experimental design to strengthen causal inference, and its primary weakness is an eschewal of effect quantification. The primary strength of the model-based framework is its rigorous attention to formal criteria for causal inference, but these criteria rely on assumptions that are difficult or impossible to test. Also, although space does not permit elaboration here, Pearl (2001, 2009, 2010, 2012, 2014; Shpitser 2013) describes a compelling framework by which causal models may be constructed using directed acyclic graphs (similar to path diagrams in SEM). Researchers within the model-based tradition are increasingly using concepts from Pearl's system.

A reasonable approach is to borrow strength from each framework. For example, a researcher might randomly assign participants to treatment and control groups (design based) and quantify the indirect effect using a definition based on potential outcomes (model based). There are active efforts to combine elements of both traditions. For example, Imai et al. (2011, 2013) describe a modified crossover design in which every participant receives both treatments, turning potential outcomes into actual outcomes. Interestingly, the latter design is similar to the within-subjects mediation design proposed by Judd et al. (2001), working within the design-based tradition. Imai et al. (2013) and Imai & Yamamoto (2013) introduce new experimental designs for scenarios where it is difficult to manipulate M directly but plausible to encourage participants to adopt certain values of M (i.e., imperfect manipulation). Such designs have the potential to strengthen internal validity without sacrificing external validity.

A contribution of the causal mediation literature has been to provide definitions of indirect effects in cases where such definitions are not always obvious (Pearl 2010). In addition, this literature has provided causally based justifications for the continued use of methods already developed in the mediation literature for more established scenarios. For example, the product of coefficients method for linear models with continuous variables, already in use for decades prior to the advent of the causal mediation literature, has since received support from the causal mediation literature when certain key assumptions have been satisfied (Imai et al. 2010a,b).

Future directions. Future work on causal mediation might take several directions. For example, most of this literature has been developed using models for binary variables. Explicit treatment of continuous variables is rare within this literature, although extensions to models using time-to-event data are considered in the next section. The degree to which the assumptions developed in the binary variable case also apply in the continuous variable case is especially important to discover, given that most applied research in psychology treats mediators and outcomes as continuous.

Second, whereas the model-based causal inference literature has focused on simple three-variable mediation models (X, M, and Y), the mainstream mediation literature within psychology routinely concerns much more complex mediation models (e.g., models with multiple mediators or multiple repeated measures of each variable). Although expansions of models used in the model-based causal inference literature have begun to appear (Albert & Nelson 2011, Imai & Yamamoto 2013, Shpitser 2013), more should be done on this front, considering the popularity of SEM in the social sciences. In addition, the necessary conditions for causal inference are underdeveloped for latent variable models, multilevel models, and mixture models.

Finally, most longitudinal mediation methods have not yet been subjected to formal causal scrutiny, with rare exceptions (e.g., Imai et al. 2011, van der Laan & Petersen 2004, VanderWeele 2010). Yet, the role of time is critical to consider in mediation analysis. For example, lag is what determines whether a potential confounder should be considered pre- or posttreatment, and some assumptions invoked in the causal mediation literature apply differently to pre- or posttreatment confounders.

Software. For any new methodological contributions to enter mainstream use, widely available, user-friendly software is a necessity. Some software for applying the ideas in the model-based causal inference literature to mediation has begun to appear—notably, recent versions of M*plus* (B. Muthén, unpublished manuscript; Muthén & Asparouhov 2014; Muthén & Muthén 1998–2014) and mediation packages/macros for R (Imai et al. 2010a, Tingley et al. 2014), Stata (Hicks & Tingley 2011), and SAS (Valeri & VanderWeele 2013). R code is provided by Coffman & Zhong (2012) for implementing marginal structural models. However, more must be done on this front for social scientists to embrace causally based mediation methods.

CATEGORICAL AND NONNORMAL VARIABLES IN MEDIATION MODELS

Although the causal inference literature has given attention to categorical mediators and/or outcomes, most treatments of mediation analysis within psychology have considered models containing only continuous variables. Many researchers may implicitly assume that any rules and procedures derived in the continuous variable case would translate directly to the discrete variable case. Sometimes this is true, but mediation analysis with categorical (or continuous but nonnormal) variables often requires special steps and is more often used in other disciplines such as medicine and epidemiology. This section provides an overview of mediation analysis with categorical independent or dependent variables and nonnormal dependent variables.

Mediation Models with Categorical Independent Variables

Binary, ordinal, or otherwise categorical independent variables in mediation models pose little difficulty. Regression methods do not require distributional assumptions about X, requiring only that X have interval properties. Ordinal or nominal X variables can be represented within a multiple linear regression framework using a series of dummy codes and treated as multiple simultaneous independent variables that represent pairwise contrasts (Hayes & Preacher 2014, MacKinnon 2008). Otherwise, no special steps are necessary. For example, if X consists of four categories, these could be represented in the form of three dummy codes, D_2, D_3, and D_4, respectively contrasting groups 2, 3, and 4 to the reference group 1:

$$\begin{aligned} Y_i &= i_2 + c'_2 D_{2i} + c'_3 D_{3i} + c'_4 D_{4i} + b M_i + e_{2i} \\ M_i &= i_3 + a_2 D_{2i} + a_3 D_{3i} + a_4 D_{4i} + e_{3i} \end{aligned} \tag{6}$$

Three relative indirect effects (Hayes & Preacher 2014) are implied by this model: $a_2 b$, $a_3 b$, and $a_4 b$. For example, $a_4 b$ represents the degree to which M is responsible for the mean difference on

Y observed between groups 1 and 4. Alternative coding systems (e.g., Helmert codes, effect codes) can be used, according to the needs of the researcher.

GLM: generalized linear model

Generalized Linear Mediation Models

Binary, ordinal, nominal, count, and time-to-event dependent variables (M or Y) are common throughout the social sciences and medical research (e.g., depression diagnosis, highest level of education attained, disease diagnosis, number of cigarettes smoked in a day, and time until death). Such outcomes are more challenging to address than are categorical predictors because mediators and outcomes are treated as endogenous variables to which distributional assumptions apply. Inclusion of such categorical dependent variables in linear regression models violates assumptions of normality and heteroscedasticity required for valid inference and can lead to out-of-range predicted values. Generalized linear models (GLMs) represent a unified family of models that can be used to accommodate such outcomes in mediation analyses (Huang et al. 2004).

In GLMs, models for each dependent variable consist of three parts:

1. a conditional response distribution: the conditional probability distribution for a dependent variable, with mean μ_i,
2. a linear predictor η_i: a linear combination of predictors, and
3. a link function $g(\mu_i)$: a function of μ_i that is linearly related to predictors in η_i.

For example, for Y and M in Equations 1 and 2, GLM uses normal (Gaussian) conditional response distributions

$$Y_i \sim N(\mu_{Yi}, \sigma_Y^2) \quad \text{and} \quad M_i \sim N(\mu_{Mi}, \sigma_M^2) \tag{7}$$

and linear predictors

$$\begin{aligned} \eta_{Yi} &= i_2 + c'X_i + bM_i \\ \eta_{Mi} &= i_3 + aX_i. \end{aligned} \tag{8}$$

Using the identity links $g(\mu_{Yi}) = \mu_{Yi}$ and $g(\mu_{Mi}) = \mu_{Mi}$ implies that

$$\mu_{Yi} = \eta_{Yi} \quad \text{and} \quad \mu_{Mi} = \eta_{Mi}. \tag{9}$$

For simple mediation models with binary M and Y, the conditional response distributions are

$$Y_i \sim Bernoulli(\mu_{Yi}) \quad \text{and} \quad M_i \sim Bernoulli(\mu_{Mi}). \tag{10}$$

The linear predictors are the same as in Equation 8 above. One possibility for the link functions are logit links, such that

$$\ln(\mu_{Yi}/(1-\mu_{Yi})) = \eta_{Yi} \quad \text{and} \quad \ln(\mu_{Mi}/(1-\mu_{Mi})) = \eta_{Mi}, \tag{11}$$

or probit links, such that

$$\Phi^{-1}(\mu_{Yi}) = \eta_{Yi} \quad \text{and} \quad \Phi^{-1}(\mu_{Mi}) = \eta_{Mi}, \tag{12}$$

depending on whether one wishes to use logistic or probit regression, respectively (Huang et al. 2004, MacKinnon & Dwyer 1993, Winship & Mare 1983). Here, Φ^{-1} denotes the inverse standard normal cumulative density function. (See sidebar Quantifying Indirect Effects.)

Using the logistic models for binary M and Y, Buis (2010), Huang et al. (2004), and VanderWeele & Vansteelandt (2010) all note that the exponentiated total, direct, and indirect effects have useful interpretations when framed as odds ratios. For example, the exponentiated indirect effect is an odds ratio expressing the factor increase in the odds of $Y = 1$ due to a one-unit increase in X occurring indirectly through M. Put another way, this odds ratio compares the odds

QUANTIFYING INDIRECT EFFECTS

There are two main ways of quantifying the indirect effect in three-variable mediation models: the product of coefficients ($a \times b$) and the difference in coefficients ($c - c'$). In the case of continuous M and Y, $a \times b = c - c'$, so it is irrelevant how the indirect effect is computed. In GLMs, the residual variance for some (e.g., binary, ordinal, Poisson) M and Y is held constant when variables are moved in and out of equations, so c cannot be directly compared to c' without rescaling to render coefficients commensurable (Bauer 2009, Buis 2010, Lockhart 2012, MacKinnon & Dwyer 1993, MacKinnon et al. 2007). Hence, it is wise to compute indirect effects using the product of coefficients method, which does not suffer from the incommensurability problem. Then the indirect effect $a \times b$ can be evaluated using traditional methods. These methods can be implemented in modern SEM software that can even combine different variable types (e.g., binary M, continuous Y) in a single model (e.g., M*plus*, Stata).

of $Y = 1$ if $X = 1$ and holding M equal to what it would have been under $X = 1$ to the odds of $Y = 1$ if $X = 1$ but holding M equal to what it would have been had $X = 0$ (VanderWeele & Vansteelandt 2010). The framework can be extended to other scenarios requiring any of a variety of response distributions and link functions.

GLM mediation models for count M or Y have received relatively little attention. One choice for the conditional response distributions would be

$$Y_i \sim Poisson(\mu_{Yi}) \quad \text{and} \quad M_i \sim Poisson(\mu_{Mi}); \tag{13}$$

the linear predictors would be the same as in Equation 8, and the link functions would be log links, such that

$$\ln(\mu_{Yi}) = \eta_{Yi} \quad \text{and} \quad \ln(\mu_{Mi}) = \eta_{Mi} \tag{14}$$

(Coxe & MacKinnon 2010, Imai et al. 2010a, Valeri & VanderWeele 2013). In the case of overdispersion (common in studies involving substance abuse or psychopathology symptoms), a two-stage zero-inflated Poisson or negative binomial model may be used instead (Wang & Albert 2012).

Mediation models for time-to-event (survival) M or Y can be specified using a Cox proportional hazards model (Roth & MacKinnon 2012), which can be cast as a special case of GLM. A common choice for the response distribution is as in Equation 10 (an indicator function denoting whether or not an event has occurred by time t), the linear predictors are as in Equation 8, and the link function is the complementary log-log link, such that

$$\ln(-\ln(1 - \mu_{Yi})) = \eta_{Yi} \quad \text{and} \quad \ln(-\ln(1 - \mu_{Mi})) = \eta_{Mi}. \tag{15}$$

For other specifications of mediation in a survival context, see discussions in Tein & MacKinnon (2003; log-survival and log-hazards models), Zhao (2012; accelerated failure time models), and Lange & Hansen (2011; Aalen additive hazards model). It is possible to incorporate right censoring (Fairchild et al. 2014, Sun 2010) and error correction (Zhao 2012). Readers interested in survival mediation models are urged to see the summary in Grotta (2012).

Mediation Models with Continuous but Nonnormal Dependent Variables

When normal theory (ordinary least squares or maximum likelihood) estimation methods are applied to continuous but nonnormal M or Y, simulation studies have shown that parameter estimates are relatively robust to nonnormality. However, standard errors (SEs) can be severely negatively biased under high nonnormality, so nonnormality-robust SEs are a recommended alternative (Finch et al. 1997; see also Pituch & Stapleton 2008). This result is not specific to mediation analysis; similar results have been shown in simulations for regression models, SEMs, and multilevel models generally, outside the mediation context. Although SEs are no longer recommended for direct use in hypothesis tests or CI construction in mediation analysis, bias in SEs signals bias in other ways of quantifying variability (e.g., Monte Carlo CIs, bootstrap CIs), so this issue must be taken seriously. MacKinnon et al. (2013) and Pituch & Stapleton (2008) make the point that bootstrapping can be used not only for addressing nonnormality of the sampling distribution of the indirect effect $a \times b$ but also for addressing nonnormality of the outcomes M and Y used in estimating this effect.

MLM: multilevel modeling

Discussion

Work on the assessment of mediation with categorical or nonnormal mediators and outcomes has progressed at a rapid pace, with most of the work on this topic having been published since 2010. GLM mediation is being increasingly addressed in the model-based causal inference mediation literature (see Huang et al. 2004). Mediation with censored outcomes and/or independent variables has been recently addressed by Wang & Zhang (2011) in a Bayesian framework.

Other kinds of interesting categorical outcome mediation models not discussed in depth here involve an alternative stage-sequential or chain reaction view of mediation for models in which at least M and Y are binary (Collins et al. 1998). Collins and colleagues define a set of criteria consistent with mediation. Also, von Eye et al. (2009) propose configural mediation analysis to investigate mediation for categorical X, M, and Y. This method involves comparing frequencies of individuals whose categorical responses fall into certain patterns that are consistent (or not) with full, partial, or no mediation. Another direction for future research likely to be of interest to psychologists involves models in which latent class variables (unobserved grouping variables) serve as predictors, mediators, or outcomes (B. Muthén, unpublished manuscript).

An alternative for fitting mediation models with nonnormal outcomes—heavy-tailed distributions or otherwise skewed dependent variables—involves robust mediation analysis based on median regression (Yuan & MacKinnon 2014). Median regression assumes neither normality nor homoscedasticity and was shown via simulation to be robust not only to departures from normality but also to outliers.

MEDIATION IN MULTILEVEL DESIGNS

The mediation models discussed to this point have been developed largely within the regression and SEM traditions, methods that implicitly assume errors to be independent and identically distributed conditional on the inclusion of relevant predictors. This assumption is violated in clustered (multilevel or hierarchical) data, which are commonly encountered in educational research (students nested in classrooms and schools), organizational research (employees nested within teams and companies), longitudinal studies (repeated measures nested within individuals), and many other settings. The main negative consequence of using traditional statistical methods with nested data is (typically) underestimation of SEs and overly narrow CIs, which lead to artificially inflated Type I error rates. Thus, it is critical to use methods that accommodate this lack of independence induced by clustering. Most studies that address mediation in such designs use multilevel modeling (MLM; also known as hierarchical linear modeling) as a modeling framework.

Multilevel Modeling Strategies

Multilevel designs are often denoted by the level at which each variable in a proposed causal sequence is assessed. For example, a design in which X is assessed at level 2 (e.g., classrooms), whereas M and Y are assessed at level 1 (e.g., students), can be described as a 2-1-1 design. A design in which there are two mediators, one assessed at level 1 and another at level 2, might be termed 2-(2,1)-1. Key advances in multilevel mediation using MLM are reviewed chronologically below.

Raudenbush & Sampson (1999) first proposed a MLM-based method to test mediation in a 2-1-1 design, and they developed a method for use in univariate MLM software. Their method involves stacking Y and M into a single column of the data set and calling this new variable R_{ijk}; each row corresponds to a level-1 unit. Simplifying their model somewhat, for instance to include only one predictor and eliminating covariates and measurement residuals, a model for R_{ijk} is:

$$R_{ijk} = D_{1ijk}Y_{jk} + D_{2ijk}M_{jk}, \tag{16}$$

where $D_{1ijk} = 1$ and $D_{2ijk} = 0$ to render $R_{ijk} = Y_{jk}$, and the reverse to yield $R_{ijk} = M_{jk}$. Functionally, Y_{jk} and M_{jk} are the level-1 variables Y and M. The level-2 model is:

$$\begin{aligned} Y_{jk} &= \pi_{Y0k} + r_{Yjk} \\ M_{jk} &= \pi_{M0k} + r_{Mjk}, \end{aligned} \tag{17}$$

where π_{Y0k} and π_{M0k} are, respectively, random intercepts for Y and M. The rs are assumed to be multivariate normal with means of zero. The random intercepts may be regressed on the level-2 X to yield the direct effect of X on M and the total effect of X on Y:

$$\begin{aligned} \pi_{Y0k} &= \gamma_{Y00} + \gamma_{Y01}X_k + u_{Y0k} \\ \pi_{M0k} &= \gamma_{M00} + \gamma_{M01}X_k + u_{M0k}. \end{aligned} \tag{18}$$

The us are assumed to be multivariate normal with means of zero. Missing from this set of equations is the effect of M on Y controlling for X. Ideally, we would want to estimate the effect of π_{M0k} on π_{Y0k} controlling for X_k (call it γ_{Y02}). MLM architecture does not permit such a model specification to be directly estimated because MLM does not allow regression relationships among random effects. To circumvent this, Raudenbush & Sampson (1999) employ a transformation to manually derive the missing parameter as a function of those already estimated. Once γ_{Y02} is obtained, the indirect effect is computed as $\gamma_{M01}\gamma_{Y02}$.

The method of Raudenbush & Sampson (1999) has strengths and limitations. It permits random coefficients, simultaneous estimation of all parameters, and unbalanced cluster sizes. Another strength of the method is that it can accommodate latent variables (but loadings must be known and supplied). Importantly, the method also recognizes that if X is a level-2 variable, then any indirect effect exerted by X must involve only level-2 variables or level-2 components of level-1 variables. However, a limitation of the method is that γ_{Y02} must be manually computed in a separate step.

Several subsequent MLM-based methods are variations on this method. Krull & MacKinnon (1999) fit a similar model to data from a 2-1-1 design, but using separate univariate multilevel models for M and Y. They extended the model to handle multiple mediators. Krull & MacKinnon (2001), Pituch & Stapleton (2011), and Card (2012) describe application of the same general method to both 2-1-1 and 1-1-1 designs, and Krull & MacKinnon (2001) also do so for 2-2-1 designs. Because a 2-2-1 design involves a 2-2 link, it cannot be handled in a straightforward manner by MLM-based methods, so a combination of single-level and MLM methods has been used.

The foregoing methods for assessing multilevel mediation using MLM all included random intercepts in the equations for M and Y, but fixed slopes. However, multilevel designs could require random slopes for 1-1 effects. Recognizing that both of the slopes relevant for mediation in a 1-1-1 design may vary randomly across clusters, Kenny et al. (2003) include random slopes for

Y regressed on X and M and for M regressed on X. The level-1 equations are:

$$\begin{aligned} Y_{ij} &= \beta_{Y0j} + \beta_{YXj} X_{ij} + \beta_{YMj} M_{ij} + \varepsilon_{Yij} \\ M_{ij} &= \beta_{M0j} + \beta_{MXj} X_{ij} + \varepsilon_{Mij}. \end{aligned} \tag{19}$$

The level-2 equations are:

$$\begin{aligned} \beta_{Y0j} &= \gamma_{Y0} + u_{Y0j} \\ \beta_{YXj} &= \gamma_{YX0} + u_{YXj} \\ \beta_{YMj} &= \gamma_{YM0} + u_{YMj} \\ \beta_{M0j} &= \gamma_{M0} + u_{M0j} \\ \beta_{MXj} &= \gamma_{MX0} + u_{MXj}. \end{aligned} \tag{20}$$

If both β_{MXj} and β_{YMj} are random, the indirect effect of X on Y is not simply $\gamma_{MX0}\gamma_{YM0}$, but rather

$$\begin{aligned} E(\beta_{MXj}\beta_{YMj}) &= E[(\gamma_{MX0} + u_{MXj})(\gamma_{YM0} + u_{YMj})] \\ &= \gamma_{MX0}\gamma_{YM0} + \tau_{MX,YM}, \end{aligned} \tag{21}$$

where $\tau_{MX,YM}$ is the level-2 covariance of these two slopes (Kenny et al. 2003). Kenny and colleagues recommend manually estimating $\tau_{MX,YM}$ with the covariance of cluster-specific slopes. Bauer et al. (2006) show how Kenny et al.'s method may be fit as a single multivariate model, permitting accurate and direct estimation of $\tau_{MX,YM}$ within the model itself. Tate & Pituch (2007) extend the model to 2-1-1 designs. Tofighi et al. (2013) point out that the covariance term $\tau_{MX,YM}$ may signal the influence of an omitted variable. This could also be the case for any covariance term, and this covariance can still be included in the indirect effect regardless of its source. Any suspected omitted variables could also be investigated.

Consistent with the majority of applications using MLM, all MLM mediation work discussed to this point treated 1-1 links as single slopes, which may be fixed or random. However, such slopes can be thought of as conflations of two slopes—one expressing a strictly within-cluster relationship and one expressing a strictly between-cluster relationship (e.g., Cronbach 1976, Davis et al. 1961, Neuhaus & Kalbfleisch 1998). In fact, 1-1 slopes are weighted averages of the "within" and "between" slopes, with the weights depending on the intraclass correlation of the predictor and on the cluster size. Cronbach (1976) termed such conflated slopes "uninterpretable blends" of slopes that differ in magnitude, meaning, and possibly even direction. It follows that any indirect effect that involves a conflated slope is itself difficult or impossible to interpret. To solve this problem, MacKinnon (2008) and Zhang et al. (2009) propose that 1-1 effects be partitioned into level-specific component effects by cluster-mean centering the predictor in each such relationship. For example, in a fixed-slope model for a 2-1-1 design, the effect of M on Y can be partitioned. The level-1 equations are:

$$\begin{aligned} Y_{ij} &= \beta_{Y0j} + \beta_{YMj}(M_{ij} - \bar{M}_{.j}) + \varepsilon_{Yij} \\ M_{ij} &= \beta_{M0j} + \varepsilon_{Mij}. \end{aligned} \tag{22}$$

The level-2 equations are:

$$\begin{aligned} \beta_{Y0j} &= \gamma_{Y0} + \gamma_{YX0} X_j + \gamma_{Y\bar{M}0}\bar{M}_{.j} + u_{YXj} \\ \beta_{YMj} &= \gamma_{YM0} \\ \beta_{M0j} &= \gamma_{M0} + \gamma_{MX0} X_j + u_{M0j}. \end{aligned} \tag{23}$$

The indirect effect in this model, $(\gamma_{MX0}\gamma_{Y\bar{M}0})$, exists only at the between-cluster level, and the 1-1 or within slope γ_{YM0} is not relevant for mediation in a 2-1-1 design. To understand why the within slope γ_{YM0} may be disregarded in a 2-1-1 design, note that within-cluster individual differences in M cannot be the result of variability in X, and hence any indirect effect that initiates with X cannot be transmitted by the within-cluster portion of M.

Similar uninterpretable blend issues arise for multiple slopes in models for 1-1-1 designs but can be addressed by centering X as well as M. In such models, there are two potential indirect effects—one operating at each level.

Multilevel structural equation modeling (MSEM): a general statistical modeling framework capable of handling clustered data, latent variables, and complex structural relationships among variables

Using the multilevel approach, the presence and extent of mediation are inferred by testing the indirect effect for significance using approaches similar to those used to test the indirect effect in single-level mediation. These approaches include a Wald test (that is, computing the SE of the indirect effect and conducting a z-test), a method based on the distribution of products, bias-corrected bootstrap CIs, or a Monte Carlo method (Krull & MacKinnon 1999, 2001; Pituch et al. 2005). Simulation studies using particular MLM mediation designs showed that CIs based on the product distribution method, bootstrap CIs, or a Monte Carlo method were generally superior, in terms of both Type I error rate and power, to selected other methods, with both normal and nonnormal data (Pituch & Stapleton 2008; Pituch et al. 2005, 2006).

Despite the flexibility of MLM for examining mediation in multilevel designs, the framework has some important limitations. As noted by Krull & MacKinnon (2001), completely MLM-based methods require that each link in the causal chain involve an outcome assessed at level 1. Preacher et al. (2010, 2011), Preacher (2011), and Card (2012) note three additional limitations of MLM, namely that (*a*) 2-1 slopes that are frequently embedded in such models are really specific to level 2 and are not cross-level slopes, (*b*) 1-1 effects are conflated by default unless the researcher explicitly separates the effects through predictor centering, and (*c*) MLM can be used to fit only a subset of models that are theoretically justifiable. Specifically, MLM is incapable of fitting models containing upward (1-2) effects, or indeed any effects terminating with a level-2 variable (Pituch et al. 2006, 2010). Even when conflated effects are decomposed into level-specific components through cluster mean centering, level-2 effects are biased when observed cluster means are used as predictors (Lüdtke et al. 2008, Preacher et al. 2010).

Multilevel Structural Equation Modeling

Several authors have described the advantages of using multilevel SEM (MSEM) for addressing all of these problems (Card 2012; Pituch & Stapleton 2011; Preacher et al. 2010, 2011). There are several strategies for combining the advantages of MLM and SEM, but the version of MSEM most often advocated and used in the mediation literature uses a model described by Muthén & Asparouhov (2009) and implemented in M*plus* (Muthén & Muthén 1998–2014). The model consists of three matrix equations, simplified here to contain only the essential terms:

$$\begin{aligned} \mathbf{Y}_{ij} &= \mathbf{\Lambda}\boldsymbol{\eta}_{ij} \\ \boldsymbol{\eta}_{ij} &= \boldsymbol{\alpha}_j + \mathbf{B}_j\boldsymbol{\eta}_{ij} + \boldsymbol{\zeta}_{ij} \\ \boldsymbol{\eta}_j &= \boldsymbol{\mu} + \boldsymbol{\beta}\boldsymbol{\eta}_j + \boldsymbol{\zeta}_j. \end{aligned} \tag{24}$$

The first two equations are, respectively, the measurement and structural models commonly used in single-level SEM. Here, $\mathbf{Y}_{ij}$ is a vector containing all measured variables, $\mathbf{\Lambda}$ is a factor-loading matrix linking the observed variables to latent components at each level, $\boldsymbol{\eta}_{ij}$ is a vector containing all latent components, $\boldsymbol{\alpha}_j$ and $\mathbf{B}_j$ contain intercepts and path coefficients (any of which may be random) linking the latent components, and $\boldsymbol{\zeta}_{ij}$ is a vector of level-1 residuals. The primary innovation of MSEM over SEM is the addition of the vector $\boldsymbol{\eta}_j$, containing all the random intercepts and slopes from the previous equation. The third equation allows for the level-2 regressions of these random coefficients on one another. This general model contains all previous models for multilevel mediation as constrained special cases (Preacher et al. 2010, 2011) and has the further benefit of providing SEM-style fit statistics in models without random slopes (Ryu & West 2009). Moreover, by representing cluster-level components of level-1 variables as

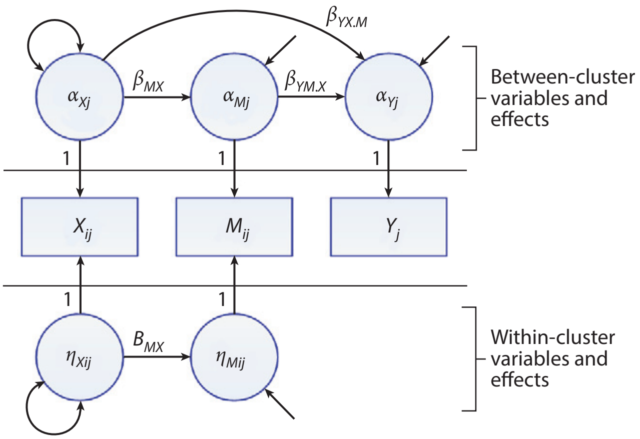

Figure 4

Multilevel structural equation model path diagram for mediation in a 1-1-2 design.

latent means (random intercepts) rather than manually computed group means, the bias noted by Lüdtke et al. (2008) is avoided. However, despite having less bias, MSEM models with more estimated parameters than MLM models sacrifice efficiency. This can (all else being equal) reduce power to detect indirect effects, particularly in the presence of low predictor intraclass correlation (Preacher et al. 2011).

An example that illustrates some of the advantages of MSEM for assessing multilevel mediation is the model in **Figure 4** for a 1-1-2 design. In this model, X and M are measured at level 1 and Y at level 2. Because Y is a strictly level-2 variable, the indirect effect must involve only the level-2 (between) components of X, M, and Y. The strictly between-cluster indirect effect is $\beta_{MX}\beta_{YM.X}$. Because the means of X and M are treated as latent variables (random intercepts), bias in the between indirect effect is avoided. Recent illustrations of MSEM for assessing mediation include Card (2012), Li & Beretvas (2013), and Tofighi & Thoemmes (2014).

Discussion

This section has covered some common MLM- and MSEM-based methods for assessing mediation in clustered data. Many extensions are possible for the methods discussed here. For example, Lockhart (2012) explores specifying MLM models for 2-2-1, 2-1-1, and 1-1-1 designs involving nonlinear $X \rightarrow M$ relationships and categorical dependent variables. Yuan & MacKinnon (2009) used Bayesian estimation to estimate the parameters of MLM models for 1-1-1 designs with WinBUGS, demonstrating some computational and inferential advantages. Ledermann & Macho (2009) propose a SEM-based "common fate" model for investigating mediation in data from individuals nested in dyads. MacKinnon (2008) describes an alternative way of assessing mediation in clustered data—adjusting the SEs for clustering and proceeding with single-level analysis (in M*plus*, invoking TYPE = COMPLEX). This adjustment probably is most appropriate when clustering is considered a nuisance and when it is not of interest to distinguish between- versus within-cluster indirect effects.

Multilevel mediation designs are, of course, not limited to two-level, fully clustered designs. For example, Pituch et al. (2010) extend MLM-based methods to accommodate a variety of mediation designs in the context of three-level data, and Preacher (2011) extends the MSEM approach to accommodate data from any three-level design. Sterba et al. (2014) and Lachowicz et al. (2014) address the assessment of mediation in partially nested designs (e.g., designs in which the treatment arm is clustered and the control arm is not). Worthwhile directions for future research would be to explore methods for assessing mediation in the presence of cross-classification (e.g., designs in which data are nested in both schools and neighborhoods; Fielding & Goldstein 2006), rolling group membership (e.g., designs featuring support groups or sports teams characterized by yearly turnover; Bauer et al. 2013), and geospatial clustering (e.g., with data nested in neighborhoods without well-defined boundaries). Finally, to date, the causal inference literature has paid only modest attention to models for multilevel data (exceptions are Raudenbush et al. 2012, VanderWeele 2010).

Partially nested design: a design in which at least one study arm is characterized by nested data and at least one study arm is nonnested

CONCLUSION

It is instructive to bear in mind Bullock et al.'s (2010) observation that "mediation is an inherently difficult subject—difficult even under favorable conditions and more difficult than the proliferation of regression-based and often-formulaic mediation analyses may suggest" (p. 555). The traditional mediation model depicted in **Figure 1** is deceptively simple and rarely appropriate by modern standards. As this survey and synthesis has shown, how to assess mediation in a given context depends on the particular theoretical model, the design of the study, the nature of the data, characteristics of the sample, and the researcher's goals—there is no universally correct approach.

The complex and diverse needs of the research community have motivated a great deal of methodological work on mediation analysis, particularly in the past decade. Methodologists have responded to these needs by developing new statistical methods tailored to specific combinations of design, model, and data. Methodologists have nearly reached consensus on some issues that were previously the subject of some debate—the importance of using longitudinal data, the need to establish support for causal inference, and methods for obtaining CIs and significance tests for indirect effects. However, there remains much active methodological research on mediation analysis. The conversation about how to best assess mediation in multilevel designs is ongoing, and the causal inference literature continues to introduce new insights into mediation analysis by addressing inferential weaknesses in more traditional approaches and by suggesting causally rigorous methods to quantify indirect effects in nonstandard situations.

Nevertheless, there remains a lack of communication between methodologists, on one hand, and psychological researchers who wish to test innovative and complex theories, on the other hand. It is hoped that this synthesis has gone some way toward bridging that divide.

DISCLOSURE STATEMENT

The author is not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

The author thanks Sonya Sterba for her valuable comments. A more extensive bibliography of articles in the four areas addressed in this review can be found at the author's website, **http://quantpsy.org**.

LITERATURE CITED

Albert JM, Nelson S. 2011. Generalized causal mediation analysis. *Biometrics* 67:1028–38

Bauer DJ. 2009. A note on comparing the estimates of models for cluster-correlated or longitudinal data with binary or ordinal outcomes. *Psychometrika* 74:97–105

Bauer DJ, Gottfredson NC, Dean D, Zucker RA. 2013. Analyzing repeated measures data on individuals nested within groups: accounting for dynamic group effects. *Psychol. Methods* 18:1–14

Bauer DJ, Preacher KJ, Gil KM. 2006. Conceptualizing and testing random indirect effects and moderated mediation in multilevel models: new procedures and recommendations. *Psychol. Methods* 11:142–63

Bentley JP. 2011. *An examination of statistical methods for longitudinal mediation modeling*. PhD thesis, Univ. Ala., Birmingham

Bollen KA, Curran PJ. 2006. *Latent Curve Models: A Structural Equation Approach*. Hoboken, NJ: Wiley

Bollen KA, Stine R. 1990. Direct and indirect effects: classical and bootstrap estimates of variability. *Sociol. Methodol.* 20:115–40

Buis ML. 2010. Direct and indirect effects in a logit model. *Stata J.* 10:11–29

Bullock JG, Green DP, Ha SE. 2010. Yes, but what's the mechanism? (Don't expect an easy answer). *J. Personal. Soc. Psychol.* 98:550–58

Campbell DT, Stanley JC. 1963. Experimental and quasi-experimental designs for research on teaching. In *Handbook of Research on Teaching*, ed. NL Gage, pp. 171–246. Chicago: Rand McNally

Card NA. 2012. Multilevel mediational analysis in the study of daily lives. In *Handbook of Research Methods for Studying Daily Life*, ed. MR Mehl, TS Conner, pp. 479–94. New York: Guilford

Cheong J. 2011. Accuracy of estimates and statistical power for testing mediation in latent growth curve modeling. *Struct. Equ. Model.* 18:195–211

Cheong J, MacKinnon DP, Khoo ST. 2003. Investigation of mediational process using parallel process latent growth curve modeling. *Struct. Equ. Model.* 10:238–62

Coffman DL. 2011. Estimating causal effects in mediation analysis using propensity scores. *Struct. Equ. Model.* 18:357–69

Coffman DL, Zhong W. 2012. Assessing mediation using marginal structural models in the presence of confounding and moderation. *Psychol. Methods* 17:642–64

Cole DA, Maxwell SE. 2003. Testing mediational models with longitudinal data: questions and tips in the use of structural equation modeling. *J. Abnorm. Psychol.* 112:558–77

Describes many key theoretical and practical issues important in longitudinal mediation analysis.

Collins LM, Graham JW, Flaherty BP. 1998. An alternative framework for defining mediation. *Multivar. Behav. Res.* 33:295–312

Coxe S, MacKinnon DP. 2010. Mediation analysis of Poisson distributed count outcomes. *Multivar. Behav. Res.* 45:1022

Cronbach LJ. 1976. *Research on Classrooms and Schools: Formulation of Questions, Design, and Analysis*. Stanford, CA: Stanford Univ. Eval. Consort.

Davis JA, Spaeth JL, Huson C. 1961. A technique for analyzing the effects of group composition. *Am. Sociol. Rev.* 26:215–25

Deboeck PR, Nicholson JS, Bergeman CS, Preacher KJ. 2013. From modeling long-term growth to short-term fluctuations: Differential equation modeling is the language of change. In *Springer Proceedings in Mathematics and Statistics:* Vol. 66. *New Developments in Quantitative Psychology*, ed. RE Millsap, LA van der Ark, DM Bolt, CM Woods, pp. 427–47. New York: Springer

Elliott MR, Raghunathan TE, Li Y. 2010. Bayesian inference for causal mediation effects using principal stratification with dichotomous mediators and outcomes. *Biostatistics* 11:353–72

Emsley R, Dunn G, White IR. 2010. Mediation and moderation of treatment effects in randomised controlled trials of complex interventions. *Stat. Methods Med. Res.* 19:237–70

Fairchild AJ, Abara WE, Gottschall AC, Tein J-Y, Prinz RJ. 2014. Improving our ability to evaluate underlying mechanisms of behavioral onset and other event occurrence outcomes: a discrete-time survival mediation model. *Eval. Health Prof.* In press

Fielding A, Goldstein H. 2006. *Cross-classified and multiple membership structures in multilevel models*. Dep. Educ. Skills, Res. Rep. RR791, Univ. Birmingham, UK

Finch JF, West SG, MacKinnon DP. 1997. Effects of sample size and nonnormality on the estimation of mediated effects in latent variable models. *Struct. Equ. Model.* 4:87–107

Frangakis CE, Rubin DB. 2002. Principal stratification in causal inference. *Biometrics* 58:21–29

Fritz MS. 2014. An exponential decay model for mediation. *Prev. Sci.* In press

Gallop R, Small DS, Lin JY, Elliott MR, Joffe M, Ten Have TR. 2009. Mediation analysis with principal stratification. *Stat. Med.* 28:1108–30

Gollob HF, Reichardt CS. 1987. Taking account of time lags in causal models. *Child Dev.* 58:80–92

Gollob HF, Reichardt CS. 1991. Interpreting and estimating indirect effects assuming time lags really matter. In *Best Methods for the Analysis of Change*, ed. LM Collins, JL Horn, pp. 243–59. Washington, DC: Am. Psychol. Assoc.

Grotta A. 2012. *Causal mediation analysis on survival data: an application on the National March Cohort*. PhD thesis, Univ. Milano-Bicocca, Milan

Gu F, Preacher KJ, Ferrer E. 2014. A state space modeling approach to mediation analysis. *J. Educ. Behav. Stat.* 39:117–43

Hafeman DM, VanderWeele TJ. 2011. Alternative assumptions for the identification of direct and indirect effects. *Epidemiology* 22:753–64

Hayes AF. 2013. *An Introduction to Mediation, Moderation, and Conditional Process Analysis: A Regression-Based Approach*. New York: Guilford

Provides a thorough and practical treatment of estimation and inference for mediation, moderation, and their combination.

Hayes AF, Preacher KJ. 2010. Quantifying and testing indirect effects in simple mediation models when the constituent paths are nonlinear. *Multivar. Behav. Res.* 45:627–60

Hayes AF, Preacher KJ. 2014. Statistical mediation analysis with a multicategorical independent variable. *Br. J. Math. Stat. Psychol.* 67:451–70

Hayes AF, Scharkow M. 2013. The relative trustworthiness of inferential tests of the indirect effect in statistical mediation analysis: Does method really matter? *Psychol. Sci.* 24:1918–27

Hicks R, Tingley D. 2011. Causal mediation analysis. *Stata J.* 11:609–15

Holland PW. 1986. Statistics and causal inference. *J. Am. Stat. Assoc.* 81:945–70

Holland PW. 1988. Causal inference, path analysis, and recursive structural equations models. *Sociol. Methodol.* 18:449–84

Huang B, Sivaganesan S, Succop P, Goodman E. 2004. Statistical assessment of mediational effects for logistic mediational models. *Stat. Med.* 23:2713–28

Describes using the generalized linear model for assessing mediation with a variety of variable types.

Imai K, Jo B, Stuart EA. 2011. Commentary: using potential outcomes to understand causal mediation analysis. *Multivar. Behav. Res.* 46:861–73

Imai K, Keele L, Tingley D. 2010a. A general approach to causal mediation analysis. *Psychol. Methods* 15:309–34

Discusses the definition, causal identification, and estimation of indirect effects under the potential outcomes framework.

Imai K, Keele L, Yamamoto T. 2010b. Identification, inference and sensitivity analysis for causal mediation effects. *Stat. Sci.* 25:51–71

Imai K, Tingley D, Yamamoto T. 2013. Experimental designs for identifying causal mechanisms. *J. R. Stat. Soc. Ser. A* 176:5–51

Imai K, Yamamoto T. 2013. Identification and sensitivity analysis for multiple causal mechanisms: revisiting evidence from framing experiments. *Polit. Anal.* 21:141–71

Jo B. 2008. Causal inference in randomized experiments with mediational processes. *Psychol. Methods* 13:314–36

Jo B, Stuart EA, MacKinnon D, Vinokur AD. 2011. The use of propensity scores in mediation analysis. *Multivar. Behav. Res.* 46:425–52

Judd CM, Kenny DA. 1981. Process analysis: estimating mediation in treatment evaluations. *Eval. Rev.* 5:602–19

Judd CM, Kenny DA, McClelland GH. 2001. Estimating and testing mediation and moderation in within-subject designs. *Psychol. Methods* 6:115–34

Kenny DA. 2008. Reflections on mediation. *Organ. Res. Methods* 11:353–58

Kenny DA, Korchmaros JD, Bolger N. 2003. Lower level mediation in multilevel models. *Psychol. Methods* 8:115–28

Krull JL, MacKinnon DP. 1999. Multilevel mediation modeling in group-based intervention studies. *Eval. Rev.* 23:418–44

Krull JL, MacKinnon DP. 2001. Multilevel modeling of individual and group level mediated effects. *Multivar. Behav. Res.* 36:249–77

Lachowicz MJ, Sterba SK, Preacher KJ. 2014. Investigating multilevel mediation with fully or partially nested data. *Group Process. Intergroup Relat.* In press

Lange T, Hansen JV. 2011. Direct and indirect effects in a survival context. *Epidemiology* 22:575–81

Lange T, Vansteelandt S, Bekaert M. 2012. A simple unified approach for estimating natural direct and indirect effects. *Am. J. Epidemiol.* 176:190–95

Ledermann T, Macho S. 2009. Mediation in dyadic data at the level of the dyads: A structural equation modeling approach. *J. Fam. Psychol.* 23:661–70

Li X, Beretvas SN. 2013. Sample size limits for estimating upper level mediation models using multilevel SEM. *Struct. Equ. Model.* 20:241–64

Lockhart LL. 2012. *Nonlinear mediation in clustered data: a nonlinear multilevel mediation model*. PhD thesis, Univ. Texas, Austin

Lüdtke O, Marsh HW, Robitzsch A, Trautwein U, Asparouhov T, Muthén B. 2008. The multilevel latent covariate model: a new, more reliable approach to group-level effects in contextual studies. *Psychol. Methods* 13:203–29

Lynch K, Cary M, Gallop R, Ten Have T. 2008. Causal mediation analyses for randomized trials. *Health Serv. Outcomes Res. Methodol.* 8:57–76

MacKinnon DP. 2008. *Introduction to Statistical Mediation Analysis*. Mahwah, NJ: Taylor & Francis

Provides a comprehensive treatment of fundamental and advanced issues in the theory and practice of mediation analysis.

MacKinnon DP, Cheong J, Pirlott AG. 2012. Statistical mediation analysis. In *APA Handbook of Research Methods in Psychology*, Vol. 2, ed. H Cooper, pp. 313–31.Washington, DC: Am. Psychol. Assoc.

MacKinnon DP, Dwyer JH. 1993. Estimating mediated effects in prevention studies. *Eval. Rev.* 17:144–58

MacKinnon DP, Fairchild AJ. 2009. Current directions in mediation analysis. *Curr. Dir. Psychol. Sci.* 18:16–20

MacKinnon DP, Fairchild AJ, Fritz MS. 2007. Mediation analysis. *Annu. Rev. Psychol.* 58:593–614

Reviews of current topics in the mediation literature.

MacKinnon DP, Kisbu-Sakarya Y, Gottschall AC. 2013a. Developments in mediation analysis. In *The Oxford Handbook of Quantitative Methods*, ed. TD Little, 2:338–60. New York: Oxford Univ. Press

MacKinnon DP, Lockhart G, Baraldi AN, Gelfand LA. 2013b. Evaluating treatment mediators and moderators. In *The Oxford Handbook of Research Strategies for Clinical Psychology*, ed. JS Comer, PC Kendall, pp. 262–86. New York: Oxford Univ. Press

MacKinnon DP, Lockwood CM, Brown CH, Wang W, Hoffman JM. 2007. The intermediate endpoint effect in logistic and probit regression. *Clin. Trials* 4:499–513

MacKinnon DP, Lockwood CM, Hoffman JM, West SG, Sheets V. 2002. A comparison of methods to test mediation and other intervening variable effects. *Psychol. Methods* 7:83–104

MacKinnon DP, Lockwood CM, Williams J. 2004. Confidence limits for the indirect effect: distribution of the product and resampling methods. *Multivar. Behav. Res.* 39:99–128

Maxwell SE, Cole DA. 2007. Bias in cross-sectional analyses of longitudinal mediation. *Psychol. Methods* 12:23–44

Maxwell SE, Cole DA, Mitchell MA. 2011. Bias in cross-sectional analyses of longitudinal mediation: partial and complete mediation under an autoregressive model. *Multivar. Behav. Res.* 46:816–41

McArdle JJ. 2001. A latent difference score approach to longitudinal dynamic structural analyses. In *Structural Equation Modeling: Present and Future*, ed. R Cudeck, S du Toit, D Sörbom, pp. 342–80. Lincolnwood, IL: Sci. Softw. Int.

Mitchell MA, Maxwell SE. 2013. A comparison of the cross-sectional and sequential designs when assessing longitudinal mediation. *Multivar. Behav. Res.* 48:301–39

Muthén B, Asparouhov T. 2009. Growth mixture modeling: analysis with non-Gaussian random effects. In *Longitudinal Data Analysis*, ed. G Fitzmaurice, M Davidian, G Verbeke, G Molenberghs, pp. 143–65. Boca Raton, FL: Chapman & Hall/CRC

Muthén B, Asparouhov T. 2014. Causal effects in mediation modeling: an introduction with applications to latent variables. *Struct. Equ. Model.* In press

Muthén LK, Muthén BO. 1998–2014. *Mplus User's Guide*. Los Angeles, CA: Muthén & Muthén. 7th ed.

Narayan AJ, Englund MM, Egeland B. 2013. Developmental timing and continuity of exposure to interparental violence and externalizing behavior as prospective predictors of dating violence. *Dev. Psychopathol.* 25:973–90

Neuhaus J, Kalbfleisch J. 1998. Between- and within-cluster covariate effects in the analysis of clustered data. *Biometrics* 54:638–45

Newland RP, Crnic KA, Cox MJ, Mills-Koonce WR, Fam. Life Proj. Key Investig. 2013. The family model stress and maternal psychological symptoms: mediated pathways from economic hardship to parenting. *J. Fam. Psychol.* 27:96–105

Pearl J. 2001. Direct and indirect effects. In *Proceedings of the Seventeenth Conference on Uncertainty in Artificial Intelligence*, pp. 411–20. San Francisco: Morgan Kaufmann

Pearl J. 2009. *Causality: Models, Reasoning, and Inference*. New York: Cambridge Univ. Press. 2nd ed.

Pearl J. 2010. The foundations of causal inference. *Sociol. Methodol.* 40:75–149

Pearl J. 2011. The mathematics of causal relations. In *Causality and Psychopathology*, ed. PE Shrout, pp. 47–65. New York: Oxford Univ. Press

Pearl J. 2012. The causal mediation formula—a guide to the assessment of pathways and mechanisms. *Prev. Sci.* 13:426–36

Pearl J. 2014. Interpretation and identification of causal mediation. *Psychol. Methods.* In press

Pituch KA, Murphy DL, Tate RL. 2010. Three-level models for indirect effects in school- and class-randomized experiments in education. *J. Exp. Educ.* 78:60–95

Pituch KA, Stapleton LM. 2008. The performance of methods to test upper-level mediation in the presence of nonnormal data. *Multivar. Behav. Res.* 43:237–67

Pituch KA, Stapleton LM. 2011. Hierarchical linear and structural equation modeling approaches to mediation analysis in randomized field experiments. In *The Sage Handbook of Innovation in Social Research Methods*, ed. M Williams, W Vogt, pp. 590–619. London: Sage

Pituch KA, Stapleton LM, Kang JY. 2006. A comparison of single sample and bootstrap methods to assess mediation in cluster randomized trials. *Multivar. Behav. Res.* 41:367–400

Pituch KA, Whittaker TA, Stapleton LM. 2005. A comparison of methods to test for mediation in multisite experiments. *Multivar. Behav. Res.* 40:1–24

Preacher KJ. 2011. Multilevel SEM strategies for evaluating mediation in three-level data. *Multivar. Behav. Res.* 46:691–731

Preacher KJ, Selig JP. 2012. Advantages of Monte Carlo confidence intervals for indirect effects. *Commun. Methods Meas.* 6:77–98

Preacher KJ, Zhang Z, Zyphur MJ. 2011. Alternative methods for assessing mediation in multilevel data: the advantages of multilevel SEM. *Struct. Equ. Model.* 18:161–82

Preacher KJ, Zyphur MJ, Zhang Z. 2010. A general multilevel SEM framework for assessing multilevel mediation. *Psychol. Methods* 15:209–33

Describes the use of multilevel structural equation modeling for multilevel mediation analysis.

Raudenbush SW, Reardon SF, Nomi T. 2012. Statistical analysis for multisite trials using instrumental variables with random coefficients. *J. Res. Educ. Eff.* 5:303–32

Raudenbush SW, Sampson R. 1999. Assessing direct and indirect effects in multilevel designs with latent variables. *Sociol. Methods Res.* 28:123–53

Richiardi L, Bellocco R, Zugna D. 2013. Mediation analysis in epidemiology: methods, interpretation and bias. *Int. J. Epidemiol.* 42:1511–19

Robins JM, Greenland S. 1992. Identifiability and exchangeability for direct and indirect effects. *Epidemiology* 3:143–55

Provides an early and influential application of Rubin's causal model to the study of mediation.

Roth DL, MacKinnon DP. 2012. Mediation analysis with longitudinal data. In *Longitudinal Data Analysis: A Practical Guide for Researchers in Aging, Health, and Social Sciences*, ed. JT Newsom, RN Jones, SM Hofer, pp. 181–216. New York: Routledge

Rubin DB. 1974. Estimating causal effects of treatments in randomized and nonrandomized studies. *J. Educ. Psychol.* 66:688–701

Rubin DB. 2004. Direct and indirect causal effects via potential outcomes. *Scand. J. Stat.* 31:161–70

Rubin DB. 2010. Reflections stimulated by the comments of Shadish 2010 and West and Thoemmes 2010. *Psychol. Methods* 15:38–46

Ryu E, West SG. 2009. Level-specific evaluation of model fit in multilevel structural equation modeling. *Struct. Equ. Model.* 16:583–601

Selig JP, Preacher KJ. 2009. Mediation models for longitudinal data in developmental research. *Res. Hum. Dev.* 6:144–64

Selig JP, Preacher KJ, Little TD. 2012. Modeling time-dependent association in longitudinal data: a lag as moderator approach. *Multivar. Behav. Res.* 47:697–716

Shadish WR. 1996. Meta-analysis and the exploration of causal mediating processes: a primer of examples, methods, and issues. *Psychol. Methods* 1:47–65

Shadish WR. 2010. Campbell and Rubin: a primer and comparison of their approaches to causal inference in field settings. *Psychol. Methods* 15:3–17

Shadish WR, Cook TD, Campbell DT. 2002. *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Boston: Houghton Mifflin

Shadish WR, Sullivan KJ. 2012. Theories of causation in psychological science. In *APA Handbook of Research Methods in Psychology*, Vol. I: *Foundations, Planning, Measures, and Psychometrics*, ed. H Cooper, pp. 23–52. Washington, DC: Am. Psychol. Assoc.

Shpitser I. 2013. Counterfactual graphical models for longitudinal mediation analysis with unobserved confounding. *Cogn. Sci.* 37:1011–35

Shrout PE, Bolger N. 2002. Mediation in experimental and nonexperimental studies: new procedures and recommendations. *Psychol. Methods* 7:422–45

Sigall H, Mills J. 1998. Measures of independent variables and mediators are useful in social psychology experiments: But are they necessary? *Personal. Soc. Psychol. Rev.* 2:218–26

Smith ER. 1982. Beliefs, attributions, and evaluations: nonhierarchical models of mediation in social cognition. *J. Personal. Soc. Psychol.* 43:248–59

Spencer SJ, Zanna MP, Fong GT. 2005. Establishing a causal chain: why experiments are often more effective than mediational analyses in examining psychological processes. *J. Personal. Soc. Psychol.* 89:845–51

Sterba SK, Preacher KJ, Forehand R, Hardcastle EJ, Cole DA, Compas BE. 2014. Structural equation modeling approaches for analyzing partially nested data. *Multivar. Behav. Res.* 49:93–118

Stolzenberg RM. 1980. The measurement and decomposition of causal effects in nonlinear and nonadditive models. *Sociol. Methodol.* 11:459–88

Stone-Romero EF, Rosopa PJ. 2008. The relative validity of inferences about mediation as a function of research design characteristics. *Organ. Res. Methods* 11:326–52

Stone-Romero EF, Rosopa PJ. 2011. Experimental tests of mediation models: prospects, problems, and some solutions. *Organ. Res. Methods* 14:631–46

Sun Y. 2010. *Methods for estimating mediation effect in survival analysis: Does weight loss mediate the undernutrition–mortality relationship in the older adults?* PhD thesis, Univ. Ala., Birmingham

Tate RL, Pituch KA. 2007. Multivariate hierarchical linear modeling in randomized field experiments. *J. Exp. Educ.* 75:317–37

Tchetgen Tchetgen EJ, Lin SH. 2012. *Robust estimation of pure/natural direct effects with mediator measurement error*. Harvard Univ. Biostat. Work. Pap. 152. **http://biostats.bepress.com/harvardbiostat/paper152**

Tein J-Y, MacKinnon DP. 2003. Estimating mediated effects with survival data. In *New Developments in Psychometrics*, ed. H Yanai, A Okada, K Shigemasu, Y Kano, JJ Meulman, pp. 405–12. Tokyo: Springer

Ten Have TR, Joffe MM, Lynch KG, Brown GK, Maisto SA, Beck AT. 2007. Causal mediation analyses with rank preserving models. *Biometrics* 63:926–34

Tingley D, Yamamoto T, Hirose K, Keele L, Imai K. 2014. Mediation: R package for causal mediation analysis. *J. Stat. Softw.* 59:5

Tofighi D, Thoemmes F. 2014. Single-level and multilevel mediation analysis. *J. Early Adolesc.* 34:93–119

Tofighi D, West SG, MacKinnon DP. 2013. Multilevel mediation analysis: the effects of omitted variables in the 1-1-1 model. *Br. J. Math. Stat. Psychol.* 66:290–307

Valeri L, VanderWeele TJ. 2013. Mediation analysis allowing for exposure-mediator interactions and causal interpretation: theoretical assumptions and implementation with SAS and SPSS macros. *Psychol. Methods* 18:137–50

van der Laan MJ, Petersen ML. 2004. *Estimation of direct and indirect causal effects in longitudinal studies*. Univ. Calif. Berkeley Div. Biostat. Work. Pap. No. 155

VanderWeele TJ. 2009. Marginal structural models for the estimation of direct and indirect effects. *Epidemiology* 20:18–26

VanderWeele TJ. 2010. Direct and indirect effects for neighborhood-based clustered and longitudinal data. *Sociol. Methods Res.* 38:515–44

VanderWeele TJ. 2011. Causal mediation analysis with survival data. *Epidemiology* 22:582–85

VanderWeele TJ, Vansteelandt S. 2009. Conceptual issues concerning mediation, interventions and composition. *Stat. Interface* 2:457–68

VanderWeele TJ, Vansteelandt S. 2010. Odds ratios for mediation analysis with a dichotomous outcome. *Am. J. Epidemiol.* 172:1339–48

Vansteelandt S, VanderWeele TJ. 2012. Natural direct and indirect effects on the exposed: effect decomposition under weaker assumptions. *Biometrics* 68:1019–27

von Eye A, Mun EY, Mair P. 2009. What carries a mediation process? Configural analysis of mediation. *Integr. Psychol. Behav. Sci.* 43:228–47

von Soest T, Hagtvet KA. 2011. Mediation analysis in a latent growth curve modeling framework. *Struct. Equ. Model.* 18:289–314

Wang L, Zhang Z. 2011. Estimating and testing mediation effects with censored data. *Struct. Equ. Model.* 18:18–34

Wang W, Albert JM. 2012. Estimation of mediation effects for zero-inflated regression models. *Stat. Med.* 31:3118–32

West SG, Thoemmes F. 2010. Campbell's and Rubin's perspectives on causal inference. *Psychol. Methods* 15:18–37

Winship C, Mare RD. 1983. Structural equations and path analysis for discrete data. *Am. J. Sociol.* 89:54–110

The first major source describing mediation analysis with discrete variables.

Yuan Y, MacKinnon DP. 2009. Bayesian mediation analysis. *Psychol. Methods* 14:301–22

Yuan Y, MacKinnon DP. 2014. Robust mediation analysis based on median regression. *Psychol. Methods* 19:1–20

Zhang Z, Zyphur MJ, Preacher KJ. 2009. Testing multilevel mediation using hierarchical linear models: problems and solutions. *Organ. Res. Methods* 12:695–719

Zhao S. 2012. *Covariate measurement error correction methods in mediation analysis with failure time data*. PhD thesis, Univ. Wash., Seattle