STATUS Submitted 20160321
REQUEST DATE 03/21/2016
SOURCE ILLiad
BORROWER BUF BF 76.5 .O94 2014 V.2
LENDERS YOM, *FDS, FXG, MRC, NIM, PAB, PGM, SST

NEED BEFORE 04/20/2016
RECEIVE DATE

TYPE Copy
OCLC # 876285608
DUE DATE

165330271

LOCAL ID
AUTHOR
TITLE The Oxford handbook of quantitative methods /
IMPRINT Oxford : Oxford University Press, 2014. ©2013

ISBN 9780199370177 (set)
FORMAT Book
EDITION
ART AU MacKinnon David
ART TITLE Developments in mediation analysis.

DATE 2013

ALERT

VERIFIED <TN:348999><ODYSSEY:
MAX COST OCLC IFM - 30.00 USD
LEND CHG
LEND RSTR

FAX NUM (716)645-3721 :ARIEL 128.205.190.99
EMAIL bufborro@buffalo.edu
AFFILIATION
COPYRIGHT US:CCG
SHIPPED
ODYSSEY 128.205.204.46/BUF
SERIES NOTE Oxford library of psychology.

BILL NOTE

BRWR NOTE From Vol. 2

LNDR NOTE

SHIP VIA Library Mail

SHIP TO
Libraries Annex
Univ at Buffalo-Doc Delivery
3850 Rensch Rd
Amherst, NY, US 14228-2712

BILL TO
same

RETURN VIA

RETURN TO

CHAPTER

16

Developments in Mediation Analysis

David P. MacKinnon, Yasemin Kisbu-Sakarya, *and* Amanda C. Gottschall

Abstract

Theories in many substantive disciplines specify the mediating mechanisms by which an antecedent variable is related to an outcome variable. In both intervention and observational research, mediation analyses are central to testing these theories because they describe how or why an effect occurs. Over the last 30 years, methods to investigate mediating processes have become more refined. The purpose of this chapter is to outline these new developments in four major areas: (1) significance testing and confidence interval estimation of the mediated effect, (2) mediation analysis in groups, (3) assumptions of and approaches to causal inference for assessing mediation, and (4) longitudinal mediation models. The best methods to test mediation relations are described, along with methods to assess mediation relations when they may differ across groups. Methods for addressing causal inference and models for assessing temporal precedence in mediation models are used to illustrate some remaining unresolved issues in mediation analysis, and several promising approaches to solving these problems are presented.

Key Words: Mediation, moderation, indirect effects, causal inference, longitudinal models, significance testing, confidence intervals

Introduction

A mediator (M) is a variable that transmits the effect of an antecedent variable (X) to an outcome variable (Y) in a causal sequence such that X causes M and M causes Y. Theories across many substantive disciplines focus on mediating processes as explanations for how and why an antecedent variable is related to an outcome variable. Intervention programs are designed to change mediating variables theorized to be causally related to the outcome variable. If an intervention program substantially changes a mediating variable that is causally related to an outcome, then a change in the mediator will produce a change in the outcome. The effect of the antecedent variable on the mediator variable is called the action theory because a manipulation is introduced to change the mediator. The effect of the mediator variable on the outcome variable is called the conceptual theory because the mediator is not directly manipulated but is theorized to affect the outcome. Mediating variables can be psychological (e.g., knowledge, beliefs, and attitudes), behavioral (e.g., interpersonal skills), or biological (e.g., serum cholesterol level). The purpose of this chapter is to review current methods used to investigate mediating variables. As comprehensive descriptions of mediation analysis are available in the literature (MacKinnon, 2008; MacKinnon, Fairchild, & Fritz, 2007), only an overview of mediation analysis is presented here. The purpose of this chapter is to supplement these resources with a description of recent advances in four major areas: (1) significance testing and confidence interval estimation of the mediated effect, (2) mediation analysis in groups, (3) assumptions of and approaches to improving causal inference for mediation, and (4) longitudinal mediation models.

History

Mediation analysis was introduced as a way to explain observed relations among variables. Modern methods for quantifying mediation models started with Wright's path analysis methods (1920, 1921). Path analysis provides a framework for the causal relations among variables and estimation of the size of those relations. About 30 years later, the elaboration method was described by Lazarsfeld and colleagues (Kendall & Lazarsfeld, 1950; Lazarsfeld, 1955) to demonstrate how an original relation between two variables would change when other variables were added into the statistical analysis. The elaboration method provided a set of analyses to investigate mediation and other third-variable effects (Hyman, 1955). Also during this time, Rozeboom (1956) described mediation as a set of functional relations where the mediator is a function of the independent variable and the dependent variable is a function of the mediator. During the '60s and '70s, models for sets of causal relations were described (Blalock, 1971; Duncan, 1966, 1975; Rosenberg, 1968), including covariance structure modeling, which combined path analysis and measurement traditions (Jöreskog, 1970; Keesling, 1972; Wiley, 1973).

Since the 1970s, many advances in methodological, statistical, and substantive aspects of mediation analysis have been made. Sobel (1982, 1986) derived the standard error for any indirect effect. Inspired by earlier work on the elaboration method (Hyman, 1955; Kenny, 2008; Mathiew, DeShon, & Bergh, 2008), Kenny and colleagues (Baron & Kenny, 1986; Judd & Kenny, 1981) described the regression equations and outlined the tests necessary for a causal steps method to assess mediation. MacKinnon and Dwyer (1993) specified the mediation regression equations, proposed different approaches to estimate the mediated effect, and provided several formulas for the standard error of the mediated effect. Applications of mediation analysis to answer substantive questions were given in social psychology (Baron & Kenny, 1986), applied psychology (James & Brett, 1984), prevention programs (MacKinnon & Dwyer, 1993), and epidemiology (Robins & Greenland, 1992). Bootstrap methods for mediation analysis were outlined (Bollen & Stine, 1993; Lockwood & MacKinnon, 1998), as were conceptualizations of longitudinal mediation (Gollob & Reichardt, 1991). A causal inference perspective on mediation was outlined by Holland (1988) for the social sciences and by Robins and Greenland (1992) for epidemiology. Recently, the similarity of mediation methods across fields such as epidemiology, medicine, psychology, and sociology have been recognized and led to several advances (MacKinnon, 2008). Since 2000, advances in mediation analysis have occurred for significance testing and confidence interval estimation (MacKinnon, Lockwood, Hoffman, West, & Sheets, 2002; MacKinnon, Lockwood, & Williams, 2004), mediation in grouped data (Krull & MacKinnon, 2001; Preacher, Zyphur, & Zhang, 2010), causal inference (Imai et al., 2010; Jo, 2008; Pearl, 2009; Sobel 2007; VanderWeele, 2008) and longitudinal data (Cheong, MacKinnon, & Khoo, 2003; Cole & Maxwell, 2003; von Eye, Mun, & Mair, 2009).

Modern Appeal

At least part of the recent interest in mediation models stems from its application to the design of interventions. The intervention is designed to change mediating variables that are hypothesized to be causally related to an outcome variable. Many drug prevention programs, for example, are designed to increase resistance skills, educate people about the risks associated with drug use, and change social norms, all of which are expected to reduce drug use. Researchers from many substantive areas have stressed the importance of assessing mediation in the evaluation of prevention and treatment studies (Baranowski, Anderson, & Carmack, 1998; Begg & Leung, 2000; Donaldson, 2001; Judd & Kenny, 1981; MacKinnon, 1994; Sandler, Wolchik, MacKinnon, Ayers, & Roosa, 1997; Weiss, 1997). The broad appeal of mediation analysis for intervention studies is evident by recommendations for its use across many fields, including nursing (Bennett, 2000, p. 419), nutrition (Kristal, Glanz, Tilley, & Li, 2000, p. 123), medicine (Begg & Leung, 2000, p. 27), and randomized clinical trials (Kraemer, Wilson, Fairburn, & Agras, 2002, p. 877).

Mediation analysis informs both action theory (the theory relating the intervention program components to the targeted mediators) and conceptual theory (the theory relating the mediating variables to the outcome variable) (Chen, 1990; Lipsey, 1993; MacKinnon, Taborga, & Morgan-Lopez, 2002), making its application to intervention studies a popular area of substantive and statistical interest

(MacKinnon, 1994). First, mediation analysis provides a way to evaluate if the intervention program has affected the mediators it was designed to change (i.e., tests of the action theory). For example, if a program is designed to change social norms, then program effects on normative measures should be statistically significant. Failure of the program to affect the mediating variables may occur because the program was ineffective or because measures of the mediating construct were inadequate. Second, mediation analysis provides a check on the conceptual theory of the program. If there is a not a significant relation of the mediating variable to the outcome, then this may indicate a failure of conceptual theory, that effects may emerge later, or that the mediator, the outcome, or both were not measured accurately. Third, an overall understanding of how the prevention program achieved, or failed to achieve, effects on the outcome can be obtained. For each mediator, tests of both action and conceptual theory provide a way to examine whether the lack of an intervention effect may result from the failure of the program to change the mediator, the failure of the mediator to change the outcome, or both. Overall, mediation analysis is useful for identifying the most effective components of a prevention program so that they can be retained or enhanced in future interventions.

Estimating the Mediated Effect

Several methods have been used to investigate mediating processes. An ideal way to study mediation is to manipulate mediators directly by randomly assigning subjects to different mediators or to investigate mediators in separate randomized studies (MacKinnon, 2008; MacKinnon, Taborga, et al., 2002; Spencer, Zanna, & Fong, 2005; West & Aiken, 1997). These designs will be discussed later in this chapter. However, these studies are impractical in many fields because manipulations often contain related components and it is inefficient to separate them, especially in the early stages of research. Even when subjects are randomly assigned to levels of a mediator, estimation of mediated effects is important, as it provides a test of whether the manipulation process worked as planned.

One widely used method to assess mediation is to measure mediating variables and outcome variables and conduct a series of statistical tests (Baron & Kenny, 1986; Kenny, Kashy, & Bolger, 1998). The first test assesses the extent to which the antecedent variable affects the outcome variable. Second, antecedent effects on the hypothesized

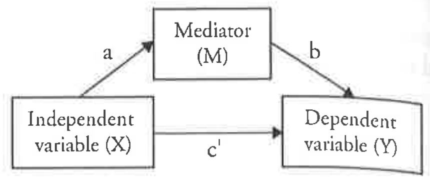

Figure 16.1 The single mediator model.

intervening or mediating variables are evaluated. Third, if effects on the outcome and the hypothesized mediating variables are substantial, then the process by which effects on the mediating variable affect changes in the outcome is assessed. More recent mediation methods have been shown to be more accurate than this causal steps method. These newer methods directly estimate the mediated effect and its standard error using the regression equations described in the next section (MacKinnon, Lockwood, et al., 2002).

Point Estimation

Difference in Coefficients Approach. In Figure 16.1 and the regression equations below, X is the antecedent variable, M is the mediator, and Y is the outcome variable. In the single mediator model (*see* Fig. 16.1), the mediated effect can be calculated in two ways (MacKinnon & Dwyer, 1993). One method, commonly used in the medical sciences, estimates the two regression equations shown below:

$$Y = i_1 + cX + e_1 \tag{1}$$

$$Y = i_2 + c'X + bM + e_2, \tag{2}$$

where i_1 and i_2 are intercepts, c is the coefficient relating the antecedent to the outcome, c' is the coefficient relating the antecedent to the outcome adjusting for the effects of the mediator, b is the coefficient relating the mediator to the dependent variable adjusting for antecedent variable, and e_1 and e_2 are unexplained variability (i.e., error). In Equation 1, the outcome variable is regressed on the antecedent variable. In Equation 2, the mediating variable is included as an additional predictor of the outcome. Using the difference in coefficients approach $(c - c')$, the mediated (i.e., indirect) effect equals the difference between c, the effect of X on Y, and c', the effect of X on Y adjusting for M. This can also be thought of as the reduction in the antecedent variable effect on the outcome variable when the mediator is included in the model. If the direct program effect coefficient c does not differ from zero when the mediator is included in the model, then

the antecedent variable effect is entirely transmitted through the mediating variable.

Product of Coefficients Approach. A second method to assess the mediated effect also involves estimation of two regression equations (Alwin & Hauser, 1975): Equation 2, presented earlier, and an additional equation shown below,

$$M = i_3 + aX + e_3, \tag{3}$$

where i_3 is an intercept, and e_3 is the unexplained variability. First, the a coefficient in Equation 3 relates the antecedent variable to the mediating variable. Second, the b coefficient in Equation 2 relates the mediator to the outcome adjusting for the antecedent variable. The product of these two coefficients (ab) is the mediated or indirect effect. The c' coefficient in Equation 2, which relates the antecedent variable to the outcome adjusting for the mediator, is the nonmediated or direct effect. The rationale behind the product of coefficients approach is that mediation depends on the extent to which the antecedent variable changes the mediator (a) and the extent to which the mediator affects the outcome variable (b). The path from the antecedent variable to the mediator to the outcome is the mediation process. The ab and $c - c'$ estimates of mediation are algebraically equivalent for ordinary least squares regression (MacKinnon, Warsi, & Dwyer, 1995). Because the ab method is easily generalized to more complicated mediation models, the product of coefficients method is recommended over the difference in coefficients method for mediation analysis.

Assumptions. For the ab estimator of the mediated effect, the model assumes that e_2 and e_3 are independent and that M and e_2 are independent (McDonald, 1997; Merrill, 1994). Other assumptions are that Equations 2 and 3 represent causal relations that are linear, additive, and recursive (James & Brett, 1984; James, Mulaik, & Brett, 2006; McDonald, 1997). Violations of this assumption can sometimes be addressed. For example, the additivity assumption implies that there is no interaction between X and M (Collins, Graham, & Flaherty, 1998; Judd & Kenny, 1981), but this interaction can be included in the model, which would allow the effect of the mediator on the outcome to differ across groups. Mediation analysis for nonrecursive models (McDonald, 1997; Sobel, 1986) and for nonlinear mediation effects (Stolzenberg, 1979) have been described in the literature.

Other important assumptions include correct causal order (e.g., X $\rightarrow$ M$\rightarrow$Y), correct causal direction (e.g., no reciprocal causation between the mediator and the dependent variable), no misspecification due to omitted variables, and minimal measurement error (Baron & Kenny, 1986; Holland, 1988; James & Brett, 1984; MacKinnon, 2008; McDonald, 1997; Pearl, 2009). Tests of sensitivity to violation of the model assumptions are the most challenging aspects of mediation analysis. As a result, the investigation of mediation processes requires a cumulative program of research using evidence from a variety of sources, including clinical observation, qualitative studies, and replication (MacKinnon, 2008).

Adding Covariates. Equations 1, 2, and 3 can be expanded to include covariates as shown below by the inclusion of variable C in Equations 4, 5, and 6. The inclusion of covariates may increase power to detect effects and may provide helpful information when investigating the sensitivity of results to different alternative explanations based on causal inference.

$$Y = i_1 + cX + dC + e_1 \tag{4}$$

$$Y = i_2 + c'X + bM + d'C + e_2 \tag{5}$$

$$M = i_3 + aX + fC + e_3 \tag{6}$$

Adding the covariate C yields a d coefficient in Equation 4 and a d' coefficient in Equation 5, corresponding to the relation of the covariate to the outcome variable controlling for the other independent variables. Similarly, the f coefficient in Equation 6 captures the relation between the covariate and M, controlling for X. For simplicity, one covariate is described here, but in practice there may be several covariates. By including one or more covariates, the values of a, b, c, and c' will likely differ from the values obtained using Equations 1, 2, and 3, where the covariate(s) are excluded from the model. Selection of covariates for a mediation model can be complicated and should typically not include posttreatment variables or colliders (*see* Pearl, 2009).

Multiple Mediators. Equations 1, 2, and 3 are easily expanded to include additional mediating variables (MacKinnon, 2000, 2008). There are separate equations for the effect of X on each mediator (Equation 3). In Equation 2, all mediators are included as predictors of Y. Multiple mediator models are postulated in several fields of research. In school-based drug prevention programs, for example, mediators such as resistance skills, social norms, attitudes about drugs, and communication skills are often targeted. The multiple mediator model more accurately reflects the multiple causes of effects on the outcome. Formulas for testing the equality of multiple mediated effects were given in MacKinnon (2000). Designs for assessing multiple mediator models in tobacco prevention research were also outlined in MacKinnon, Taborga, et al. (2002) and West and Aiken (1997). The multiple mediator model allows for positive and negative mediated effects, called inconsistent models (Blalock, 1969; Davis, 1985). These inconsistent models and their relation to suppression and confounding effects are described in MacKinnon, Krull, and Lockwood (2000). An inconsistent mediation model can occur when an intervention component does not work as planned, leading to a change in a mediator that actually increases problem behavior, whereas other components of the program reduce the problem behavior. In a study of the mediating mechanisms in a program to prevent anabolic steroid use among high school football players, the program increased knowledge of the reasons to use anabolic steroids, which actually increased intentions to use steroids in the future (MacKinnon, Goldberg, et al., 2001). Fortunately, there were other mediational processes that led to an overall beneficial effect of the intervention program (i.e., reduced intentions to use steroids in the future). These types of counterproductive mediators can be identified with multiple mediator models, which improves future programs by avoiding these iatrogenic effects.

Another type of multiple mediator model has several mediators in a sequence. In a three-path mediation model, an independent variable is hypothesized to affect one mediator, which affects a second mediator, which, in turn, affects an outcome (i.e., program $\rightarrow$ mediator1 $\rightarrow$ mediator2 $\rightarrow$ outcome). Taylor, MacKinnon, and Tein (2008) evaluated several tests of mediation for the three-path mediation model and found that a joint significance test of each of the three paths, as well as resampling methods, were the best tests of this more complicated mediation relation.

Models with multiple antecedent variables, mediators, and outcomes require structural equation modeling to more accurately estimate the relevant parameters. Matrix equations are used to organize the many parameters for these models and estimate the many mediated effects. Mediated (i.e., indirect) effects are commonly included in comprehensive structural equation models (Bollen, 1987; MacKinnon, 2008; Sobel, 1982).

Standard Error

Confidence intervals and significance tests require a measure of the estimate's variability. The standard error of the product of two random variables, a and b, is $\sigma_{ab} = (\sigma_a^2 b^2 + \sigma_b^2 a^2 + \sigma_a^2 \sigma_b^2)^{1/2}$ (Goodman, 1960, 1962). Alternatively, the multivariate delta standard error (Sobel, 1982) does not include the $\sigma_a^2 \sigma_b^2$ term (Baron & Kenny, 1986; MacKinnon & Dwyer, 1993). The standard error of the ab mediated effect obtained using either method has low relative bias in sample sizes of at least 50 for the single mediator model when the data are normally distributed (MacKinnon et al., 1995; MacKinnon, Lockwood, et al., 2002). As summarized by MacKinnon, Lockwood, et al. (2002), there are also formulas for the standard error of $c - c'$, such as $(\sigma_c^2 + \sigma_{c'}^2 - 2r\sigma_c\sigma_{c'})^{1/2}$, where $r\sigma_c\sigma_{c'}$ is the covariance between c and c' (McGuigan & Langholtz, 1988; Clogg, Petkova, & Shihadeh, 1992; Freedman & Schatzkin, 1992). The standard error can be used to compute confidence limits for the mediated effect for $c - c'$. Confidence intervals have been widely recommended for reporting research results because they enable researchers to consider the size of an effect as well as its statistical significance (Harlow, Mulaik, & Steiger, 1997; Krantz, 1999).

Significance Testing and Confidence Interval Estimation of the Mediated Effect

There have been extensive statistical and methodological developments in significance testing and confidence interval estimation of mediated effects during the last 25 years. Significance tests are typically based on normal theory. The product of two normally distributed random variables, such as the ab product estimator of the mediated effect, is not normally distributed (Craig, 1936; Springer & Thompson, 1970). Rather, the distribution of the product of two normally distributed random variables is a complicated function (Springer, 1979) that is normal only in special cases. Consider two standard normal random variables with a mean of 0—their product will have a kurtosis of six (Meeker, Cornwell, & Aroian, 1981). As a result, significance tests based on normal theory are not appropriate for testing an estimate of the mediated effect. Simulation studies have demonstrated that the statistical power and Type I error rates are too low for most tests of significance based on normal theory (i.e., using the ratio of the mediated effect to its standard error) (MacKinnon, Lockwood, et al., 2002). In one study examining the necessary sample size to detect

a mediated effect, Fritz and MacKinnon (2007) showed that tests for mediation based on normal theory or causal steps (Baron & Kenny, 1986) have larger sample size requirements than either bootstrapping methods or tests based on the distribution of the product because the latter methods account for the non-normal distribution of the product.

Because the ab-mediated effect is not normally distributed, confidence limits based on normal theory are imbalanced (Fritz & MacKinnon, 2007; MacKinnon, Lockwood, et al., 2004; MacKinnon et al., 1995). MacKinnon and colleagues (MacKinnon, Fritz, Williams, & Lockwood, 2007) have described a computer program that computes critical values for confidence limits based on the distribution of the product. Resampling methods such as bootstrapping also provide more accurate confidence limits for the mediated effect because they accommodate the non-normal distribution of the product (Arbuckle, 1997; Bentler, 1995; Hall, 1992; Jöreskog & Sorbom, 2001). MacKinnon, Lockwood, et al. (2004) and Williams and MacKinnon (2008) compared single-sample methods (including the distribution of the product) to several computer-intensive methods and found that the distribution of the product and bootstrap sampling methods led to the most accurate confidence limits. Although the bias-corrected bootstrap appeared to have the most accurate confidence limits, there was some evidence that for small sample sizes and small mediated effects there were excess Type I error rates (e.g., 0.07 instead of 0.05). These inflated Type I error rates do not appear to occur for the distribution of the product and are reduced in larger sample sizes (Fritz, Taylor, & Mackinnon, 2010).

Bayesian Methods

One promising new approach to mediation analyses is based on Bayesian mediation methods (Yuan & MacKinnon, 2009), which can complement the frequentist approaches to mediation analysis described in this chapter by providing an alternative framework to investigate mediation. Bayesian methods allow for the incorporation of prior information about mediation relations such as the size of the relation or the distribution of relevant variables. Bayesian methods may be especially useful for studies with small sample sizes in research areas where there is considerable prior information from other studies that can be incorporated into the statistical analysis. Although Bayesian methods may require changes in the approach to data analysis, they allow for straightforward applications to both complex and simple mediation models. Bayesian approaches may also provide a natural way to investigate complicated assumptions regarding causal mediation (Elliott, Raghunathan, & Li, 2010). The application of Bayesian versus frequentist statistical approaches has generated some controversy (Little, 2006) but there is plenty of room for Bayesian and frequentist methods to provide complementary approaches to the scientific investigation of mediation.

Effect Size Measures

In addition to the statistical significance of a mediated effect, recent research has addressed the need for mediation effect size measures that provide a practical and intuitive understanding of the effect. For individual paths in the mediation model, standardized regression coefficients are useful measures of effect size because they are partial correlations (MacKinnon, 2008). For example, a researcher could state that 10% of the variance in the mediator is explained by the antecedent variable. For the mediated effect, the proportion of the total effect that results from the indirect effect, also called the proportion mediated $(1 - (c'/c) = ab/(ab + c'))$, and the ratio of the indirect effect to the direct effect (ab/c') are measures of effect size. For example, a researcher could state that 30% of the antecedent effect on the outcome variable was associated with the mediating variable. Simulation studies have found that the proportion mediated requires a sample size of at least 500 to stabilize unless the effects are large, particularly the direct effect c' (Freedman, 2001; MacKinnon et al., 1995). For the ratio of the indirect effect to the direct effect, larger sample sizes of close to 1000 are required. It is important to emphasize that the accuracy of these methods is also a function of population effect size. Finally, the R^2 effect size measure focuses on the amount of variance in Y explained by both X and M, and this method works well for many different mediation models (Fairchild, MacKinnon, Taborga, & Taylor, 2009). One promising effect size measure is the mediated effect divided by the standard deviation of the outcome variable; this scales the mediated effect in standard deviation units of the outcome variable.

Categorical and Count Outcomes

In many studies, the dependent variable is categorical (e.g., whether a person used drugs or not) or a count (e.g., the number of times an event occurred). In the binary categorical case, Equations 1 and 2 are estimated with logistic or probit regression. Logistic and probit regression coefficients can be distorted because not only are the coefficients a function of the true relations between variables, but they are also a function of a fixed error term in each regression equation. The $c-c'$ method of estimating mediation can be distorted in these models because the parameter estimate of c' depends on both the indirect effect and the scaling of Y in Equation 2 (MacKinnon & Dwyer, 1993). Because the $c-c'$ estimate can be incorrect when the outcome is categorical, it is no longer equivalent to the ab estimate. One method to put the $c-c'$ estimate in the same metric as the ab estimate is to standardize the regression coefficients prior to estimating the mediated effect (Winship & Mare, 1983). With standardization of the coefficients, $c-c'$ is close to ab (MacKinnon & Dwyer, 1993; MacKinnon, Taylor, Yoon, & Lockwood, 2009). However, it is usually best to use the product of coefficients approach, ab, in logistic and probit regression mediation analysis because it is more general and does not require standardization. To construct confidence intervals, the distribution of the product and resampling methods are best for mediation estimation with categorical data.

An extension of the mediation model to the generalized linear model framework, which accommodates both logistic and probit regression as well as other methods such as survival analysis, has been outlined (MacKinnon, Lockwood, Brown, Wang, & Hoffman, 2007). In a preliminary study of mediation analysis of count outcomes using Poisson regression models, it was found that ab and $c-c'$ were usually comparable (Coxe & MacKinnon, 2010). Like the logistic and probit regression models, ab does not equal $c-c'$ when the outcome is time to an event (e.g., death) and survival analysis is used (Tein & MacKinnon, 2003). Tein and MacKinnon (2003) investigated the proportional hazards regression procedure, which allows for covariates to be incorporated into the survival model. The study found that the differences between ab and $c-c'$ decreased as the sample size increased; however, the standard errors of ab and $c-c'$ were nearly identical in sample sizes of at least 100. Recently, Pearl (2010b) has proposed a *Mediation Formula*, which presents a general approach to estimating mediation for both parametric and nonparametric relations among categorical variables. An interesting aspect of this method is that the difference score (i.e., $c-c'$) and the product of coefficients (i.e., ab) methods represent two unique ways to approach mediation, and each approach may be appropriate in different contexts.

Non-Normality

Research on non-normality and mediation analysis (Bollen & Stine, 1993, Finch, West, & MacKinnon, 1997; Lockwood, 2000) suggests the use of resampling approaches to mediation analysis, as these methods make fewer assumptions about the underlying distribution of the indirect effect (Bollen & Stine, 1990; Manly, 1997; Noreen, 1989). Computer programs have been written to conduct bootstrap estimation of the mediated effect (Lockwood & MacKinnon, 1998; Preacher & Hayes, 2004) and extended to include more computer-intensive tests (MacKinnon et al., 2004; Williams & MacKinnon, 2002). Resampling methods are a method of choice to investigate models with non-normal distributions (Williams & MacKinnon, 2008; Cumming & Finch, 2001). Recent approaches relax assumptions about the distribution of the variables involved in mediation analysis, including variables that may be affected by outliers (Zu & Yuan, 2010); when data are not normally distributed, nonparametric methods tend to outperform parametric methods.

Small Samples

When one of the two paths in the mediated effect is small, tests of the mediated effect have low power, and the confidence interval tends to be too wide to detect the mediated effect for sample sizes less than 400 (Fritz & MacKinnon, 2007). Many important studies have sample sizes ranging from 30 to 200, such as studies examining a specific population (e.g., autistic children, high risk groups, burn victims, persons at a drug treatment clinic), using very expensive measurement (e.g., magnetic resonance imaging, positron emission tomography), or collecting very intensive repeated measurements (e.g., daily assessments). Several methods have been proposed for increasing power in research studies with limited sample sizes (Hoyle, 1999) by improving the research design, measurement, or analysis.

Design. Power to detect the mediated effect can be increased through the design of a study by using extreme groups or planned missing data. Extreme group analyses are conducted to enlarge the effect size of an intervention in situations where it is possible to select groups at extreme levels of the mediator or outcome. Because extreme groups are selected, the corresponding effect size will theoretically be larger. Given a fixed sample size, researchers select participants based on scores on the mediator or the outcome variable to maximize statistical power (Alf & Abrahams, 1975). One example is the case of selecting persons for a research study that are the lowest on a mediator targeted by the intervention (Pillow, Sandler, Braver, Wolchik, & Gersten, 1991; West, Sandler, Baca, Pillow, & Gersten, 1991). In this approach, the size of the a path relating the intervention to the mediator is maximized. Similarly, participants could be selected based on their level of the dependent variable at baseline. An alternative is to oversample extreme cases based on principles of optimal design to detect effects (McClelland & Judd, 1993) and to obtain one or more pretest measures to assess regression to the mean (Pitts, 1997). Although the extreme group method for increasing power to detect a mediation effect shows promise, there are limitations (Preacher, Rucker, MacCallum, & Nicewander, 2005). One limitation to the extreme group method is the phenomenon of regression to the mean, whereby extremes at one time will tend to score closer to the mean at a later time.

When high-quality but expensive (or time-consuming, difficult, or labor-intensive) measurements are required, power can be improved by using a planned missing data design (Graham, Hofer, & MacKinnon, 1996; Graham, Hofer, & Piccinin, 1994; Graham, Taylor, & Cumsille, 2001) where the expensive measurement is obtained for only a subsample of participants. Here the subsample of participants who were only administered the inexpensive measure(s) can provide some additional power when incorporated into an analysis that uses the expensive measurement (Enders, 2010; Schafer, 1997). The power to detect mediation with small sample sizes is likely influenced by the amount of missing data and the number of measured variables and is an important area of investigation (Graham, Hofer, Donaldson, MacKinnon, & Schafer, 1997).

Measurement. Collecting baseline measures of variables in the mediation model can reduce unexplained variability, thereby increasing the ability to detect effects. To the extent that the baseline measures are strongly related to the other measures, there will be an increase in statistical power. Another way to increase power is with improved measurement of the constructs, which can reduce measurement error and increase statistical power. Even when the reliability of variables in a mediation model is close to 1, it may be possible to increase power to detect mediation relations by purifying measures so that they have greater validity. To illustrate, assume that norms are an important mediator of an intervention effect on an outcome variable. It may be possible to increase power to detect a mediated effect in this intervention study by measuring the most potent part of norms that leads to the mediation relation. For example, norms among friends may be a more valid mediator of the intervention effect and the use of this mediator, rather than the general norm mediator, will increase power to detect mediated effects. As described elsewhere (MacKinnon, 2008), one way to view mediation analysis is as a measurement task, where understanding and evidence of a true mediator accumulates as the measure of the mediator is improved.

Analysis. Bayesian methods (Yuan & MacKinnon, 2009) can be applied to mediation analysis when prior information on the parameter values is available. Prior information is incorporated into the parameter estimates, and this is especially useful when the sample size is small and effects are underpowered. Permutation tests may also be ideal for small sample sizes because they use the available data to construct all possible data sets that could have been observed. Permutation tests for mediation have been outlined (MacKinnon, 2008) but not yet thoroughly investigated in small sample sizes (MacKinnon & Lockwood, 2001; Taylor & MacKinnon, 2012). Early work on the permutation test suggests an elevated Type I error rate when one of the two paths in the mediation relation is zero and the other path is non-zero.

Mediation Analysis in Groups

Models with moderation and mediation are important because they provide information about how X affects Y and for which groups of participants X affects Y. Models with moderation and mediation can be used to investigate the extent to which individual paths differ across subgroups, whether a mediated effect differs across subgroups, and whether a moderator effect is explained by a mediator. In moderation of a mediated effect, the mediated effect differs for subgroups of participants (James & Brett, 1984; Judd, Kenny, & McClelland, 2001). For example, mediated effects may differ across cohorts (MacKinnon, Cheong, Goldberg, Williams, & Moe, 2002), age groups, gender, or ethnic groups. A statistical test for the equivalence of the mediated effect and tests of the equality of a, b, and c' also provide information about whether the action theory (i.e., how the program will change mediators) holds across subgroups (i.e., is invariant) and whether the conceptual theory (i.e., how mediators are related to the outcome) holds across subgroups (MacKinnon, 2008). Moderation of a mediated effect is more complex when the moderator variable is continuous, and researchers often categorize the continuous variable but categorization may lead to loss of information (MacCallum, Zhang, Preacher, & Rucker, 2002).

In mediation of a moderator effect, the mediator is intermediate in the causal sequence from an interaction effect to a dependent variable. The purpose of mediation of a moderator effect is to determine whether a mediating variable explains the interaction effect because the mediator transmits the effect of the interaction to the outcome variable. For example, assume there is an interaction effect between program exposure and self-esteem such that program effects on drug use differ as a function of self-esteem. The program effect that changes as a function of self-esteem leads to changes in the resistance skills mediator, which then leads to reduced drug use. Morgan-Lopez and MacKinnon (2001) and MacKinnon (2008) have described an estimator of the mediated moderator effect, but this proposed estimator requires further development and evaluation.

In treatment and prevention studies, the examination of mediation in program effects addresses *how* the program achieves its effects (Donaldson, 2001; MacKinnon, 2001; MacKinnon & Dwyer, 1993; MacKinnon, Weber, & Pentz, 1989; MTA Cooperative Group, 1999; Sandler et al., 1997). Moderator variables address *for whom* the intervention is effective (Aiken & West, 1991; Baron & Kenny, 1986). Models with both mediators and moderators allow for the simultaneous assessment of how a program works (mediation) and whether the program works differentially for groups of participants (moderation). These models represent an attempt to incorporate multiple elements of the program design into a single analysis for program evaluation (Lipsey, 1993; Sidani & Sechrest, 1999; West & Aiken, 1997).

A General Model of Moderation and Mediation

Early strategies to test models that combine moderator effects (i.e., interactions) and mediator effects tended to focus on estimating the two effects in separate situations (Baron & Kenny, 1986; James & Brett, 1984; Merrill, 1994; Merrill, MacKinnon, & Mayer, 2006; Morgan-Lopez, 2002; Morgan-Lopez, Castro, Chassin, & MacKinnon, 2003; Morgan-Lopez & MacKinnon, 2001). Recently, comprehensive models with both moderation and mediation have been described (Edwards & Lambert, 2007; Fairchild & MacKinnon, 2009; Hoyle & Robinson, 2004; MacKinnon, 2008; Preacher, Rucker, & Hayes, 2007) but require more evaluation. The equations below describe a combined moderator and mediator model for a single mediator and a single moderator (Fairchild & MacKinnon, 2009; MacKinnon, 2008). The model includes moderation of a mediated effect, mediation of a moderator effect, and other types of moderator and mediator effects as special cases of the model. The single mediator model is extended in several ways. First, the moderator variable, Z, is added along with its interaction with X, represented by XZ. Second, the interaction of the moderator with M is represented by MZ. In this model, a test of homogeneous action theory is included in the test of the a_3 coefficient; this tests whether the a parameter differs across the Z groups. A test of homogeneous conceptual theory is included in the test of the b_2 coefficient; this tests whether the b parameter differs across the Z groups. The test of whether the c' parameter is different across the Z groups is obtained by testing the c'_3 coefficient. The results from these tests are algebraically equivalent to the model where effects are estimated separately in each group when there are no higher-order interactions, heterogeneous variability, and functional form across two groups (MacKinnon, 2008).

$$Y = i_1 + c_1X + c_2Z + c_3XZ + e_1 \qquad (7)$$

$$Y = i_2 + c'_1X + c'_2Z + c'_3XZ + b_1M + b_2MZ + e_2 \qquad (8)$$

$$M = i_3 + a_1X + a_2Z + a_3XZ + e_3 \qquad (9)$$

Moderation of a mediated effect is tested by taking the difference between the mediated effects from each group and dividing this difference by its standard error (MacKinnon, 2008). If there are more than two groups, then contrasts between mediated effects for pairs of groups can be tested if the number of groups is low. If the moderator is continuous, a statistical test of the moderation of a mediated effect is more complex.

The above model can be made even more general by adding two more interaction terms, thereby allowing for testing of more complicated forms of moderation and mediation. The XM and XMZ interactions can be added to Equation 8 to form a more general model (MacKinnon, 2008). These interactions test whether the relation of M to Y differs across levels of X and Z. Often these relations are assumed to not exist and the terms are omitted from the model. In general, these models test mediation effects conditional on one or more moderator variables.

Point estimates and standard errors for mediation of a moderator effect in the single mediator model are accurate at sample sizes as small as 100 (Merrill et al., 2006). However, the standard errors can be too small with non-zero direct effects (Morgan-Lopez & MacKinnon, 2001, 2006). A non-zero direct effect implies that after including the mediator in the model, there is a residual relation between the independent variable and the outcome. It is common in social science research to have mediators that do not fully explain the relation between the independent variable and the outcome; this is called partial mediation. Morgan-Lopez (2003) also explored Type I error rates and empirical power of four different methods to detect simple mediation effects in a mediated moderation model. This work showed that the method using asymmetric confidence intervals had Type I error rates closest to the nominal Type I error rate of 0.05. The bias-corrected bootstrap method was the most powerful method overall, although the differences between the bias-corrected bootstrap and the asymmetric confidence interval methods were minimal. Because the bias-corrected bootstrap had elevated Type I error rates in some circumstances, the overall recommendation is to use the asymmetric confidence interval method or the percentile bootstrap. Fairchild (2008) examined the performance of several tests for moderation of a mediated effect using interaction effect sizes typically found in the literature (i.e., explaining 1%–3% of the total variance). When the interaction effect explained 1% of the overall variance, none of the tests had 0.8 power to detect effects for sample sizes less than 1000. When the interaction effect explained 3% of the overall variance, tests had at least 0.8 power to detect effects for sample sizes of at least 300. The distribution of the product test had greater power to detect interaction effects than either the multivariate δ test or the test of joint significance.

Multilevel Mediation

Multilevel mediation models also include grouping of participants, but the grouping is determined by sampling characteristics such as one's school, clinic, or family. Individual observations are not independent of the other observations within the same group (i.e., cluster). Thus, if clustered data is evaluated at the individual level, ignoring the multilevel data structure, then the assumption of independent observations is violated. This violation may lead to underestimated standard errors of estimates, which result in an inflated Type I error rate (Krull & Mackinnon, 1999, 2001). Violating the assumption of independent observations can be resolved by incorporating multilevel analysis into mediation analysis. In addition, conducting mediation analysis in the multilevel framework allows researchers to investigate mediation effects at different levels and reach more detailed conclusions about mediating mechanisms (Hofmann & Gavin, 1998). For example, when evaluating the mediating mechanisms of a prevention program delivered at the school level, investigators may find that the mediating mechanisms working at the school level are different from the mechanisms working at the individual level. At the school level, overall drug norms may mediate the program effect on the individual student's drug use, whereas an individual's resistance skills may work as a mediator at the individual level. Mediation methods for multilevel data (e.g., data from schools, clinics, families) have been described (Krull & MacKinnon, 1999, 2001; MacKinnon, 2008; Raudenbush & Sampson, 1999) and evaluation of these models has been conducted in simulation studies (Krull & MacKinnon, 1999, 2001; Pituch, Stapleton, & Kang, 2006).

There are several options for specifying mediation effects in multilevel modeling, depending on whether the independent variable, mediator, and/or the dependent variables are at the individual or group (cluster) level (Krull & MacKinnon, 2001). In some situations, the independent variable represents group-level characteristics, whereas the mediator and the dependent variable represent individual level characteristics. For example, one might hypothesize that style of the team leader (X at the team level) may affect individual employee's job satisfaction (Y at the employee level) indirectly by affecting the employee's perceived autonomy (M at the employee level). In other situations, the independent and mediating variables are group-level variables, whereas the dependent variables are individual-level variables, such as the relation between the class size and individual student's achievement mediated by teacher's level of fatigue. Equations 7, 8, and 9 can be written at the individual and group level depending on the levels of the independent, mediating, and dependent variables (*see*

Krull & MacKinnon, 1999; MacKinnon, 2008). Consequently, the path coefficients a and b are estimated at the individual or group level.

Another flexibility of multilevel mediation is that one or both of the a and b path coefficients can be modeled as fixed or random effects. In other words, the effect of the independent variable on the mediator and/or the effect of the mediator on the dependent variable can be treated as constant (i.e., fixed) or varying (i.e., random) across clusters. For example, in a study of group therapy treatments (e.g., behavior focused vs. cognition focused) on patient's social anxiety, researchers may be interested in the mediating mechanism of social interaction skills. Patients are clustered into therapy groups, and the treatment is administered at the cluster level (not at the individual level). In such cases, the path coefficient b may vary across therapy groups (i.e., clusters), and, therefore, no single values of b can be applied to all therapy groups. In other cases, both the a and b path coefficients vary significantly across groups, which can result in a non-zero covariance between the a and b paths ($\sigma_{ab} \neq 0$). Thus, the covariance of a and b, σ_{ab}, is taken into account to estimate standard errors and point estimates of the mediated effects (Kenny et al., 1998). There are several ways to estimate the covariance between the random coefficients a and b. Kenny et al. (1998) used resampling methods to estimate the covariance between a and b. Bauer, Preacher, and Gil (2006) have demonstrated that this covariance can be calculated by combining Equations 2 and 3 into the same analysis and directly estimating the covariance among the random effects, as is possible in SEM software programs such as Mplus (Muthén & Muthén, 2001). More recent developments provide multilevel approaches (Zhang, Zyphur, & Preacher, 2009) and a comprehensive structural equation model that incorporates the multilevel structure of the data (Preacher, Zyphur, et al., 2010). Mediation models for data collected from dyads (i.e., clusters with only two members, such as husband–wife and mother–child pairs) have also been outlined (Dagne, Brown, & Howe, 2007; Ledermann & Macho, 2009).

Causal Inference in Mediation

Recently, promising approaches to improve causal inference in mediation analysis have been proposed (Frangakis & Rubin, 2002; Holland, 1988; Jo, 2008; Kaufman, MacLehose, & Kaufman, 2004, Murphy, Van der Laan, Robins, & Conduct Problems Prevention Research Group, 2001; Pearl, 2009, 2010a; Robins & Greenland, 1992; Robins, Mark, & Newey, 1992; Rubin, 2004; Shipley, 2000; Sobel, 1998a, 1998b, 2007; Winship & Morgan, 1999) but have not been systematically evaluated in simulation studies and applied settings. These new models primarily address the effects of omitted variables on a mediation analysis (MacKinnon, 2008). In the single-mediator model, bias introduced by omitted variables occurs for the X to M and M to Y relations. Randomization of X balances the omitted variables to experimental conditions and reduces bias in the X to M relation. Holland (1988), who first noted the potential bias in the relation of M to Y, applied Rubin's (1974) causal model to mediation and showed that under some assumptions (especially random assignment), the typical regression coefficient for the effect of randomized X on Y, c, and the randomized effect of X on M, a, are valid estimators of the true causal effect (*see also* Sobel, 2007, for a recent extension of this approach). The regression coefficient of M on Y, b, is not an accurate estimator of the causal effect because this relation is correlational. So even when X is randomized, alternative explanations exist for mediation because M is not randomly assigned but, rather, is self-selected by participants. The estimator, c', is also not an accurate causal estimator of the direct effect controlling for M. The limitations of interpreting the b coefficient have been described in different substantive areas (Robins & Greenland, 1992; Winship & Morgan, 1999).

Sequential Ignorability Assumption

The problems raised by Holland (1988) in the causal interpretation of the single-mediator model are more clearly specified in the sequential ignorability assumption (Imai, 2010; Lynch, Cary, Gallop, & ten Have, 2008; ten Have et al., 2007). There are two parts of the sequential ignorability assumption. Sequential Ignorability A assumes that the relation of X to M is not affected by other variables. Randomization of participants to levels of X ensures that this assumption is satisfied asymptotically. The many possible omitted variables from the X to M relation are ignorable because randomization of X makes all other variables equivalent between the levels of X. Sequential Ignorability B requires that M is ignorable in its relation to Y. It is assumed that M is randomly assigned to participants at each level of X. That is, at all levels of M the values of Y are unrelated to other variables because of the assumed random assignment of persons to the levels of M. In reality, M is not randomly assigned as the participant usually self-selects their value of M. Using a treatment versus control design, Sequential Ignorability B, assumes that both X and M are randomly assigned. However, the values of M are not randomly assigned, and there are likely differences in the M to Y relation for the same participant if in the control condition versus in the treatment condition. It is even possible that the same participant in the control condition would have an M–Y relation that is opposite to the M–Y relation had they been in the treatment condition. Sequential Ignorability B is a difficult assumption to satisfy, but there are several methods to address this issue.

The extent to which sequential ignorability is a valid assumption may differ depending on the type of mediating variable. In intervention research, the mediators are selected because theory and prior empirical research suggest that they are causally related to the outcome variable. As a result, the b effect is often considered to be known and merely requires that the levels of M be changed. In this case, the manipulation that changes the X–M relation will have the same expected change in the M–Y relation. However, it may be possible that the relation of M to Y is not completely causal. Researchers can address the sequential ignorability assumption with theoretical consideration of the mediators measured in the study and the extent to which both X and M can be considered randomized across participants by the treatment assignment. For example, if dose is the mediator and administration of pills is the intervention, then it is likely that there is a monotonic relation between M and Y, and therefore Sequential Ignorability B may be a valid assumption. If the intervention is encouragement to take pills, then the M–Y relation may be more complicated because of the participant's choice in their exposure to M, and Sequential Ignorability B may be less likely to hold.

Sensitivity Analysis. Sensitivity analysis is a way to assess the influence of omitted variables on the observed mediation relations, including the relation of M to Y. The goal of these methods is to assess how large a confounder effect (i.e., Sequential Ignorability B) on the M–Y relation must be to invalidate conclusions about mediation (Frank, 2000; Li, Bienias, & Bennett, 2007; Lin, Psaty, & Kronmal, 1998; Rosenbaum, 2002). One way to conduct sensitivity analyses for mediation would be to systematically increase the correlation between the errors in Equations 2 and 3 and evaluate how much the coefficients change. The covariance between the errors in Equations 2 and 3 reflects the contribution of omitted variables to the observed relation of M on Y. When one or both assumptions of Sequential Ignorabilty have been violated, VanderWeele (2008, 2010) has formalized several useful methods to probe bias in mediation relations based on the marginal structural model.

Instrumental Variable Methods. Instrumental variable methods provide another way to address the influence of omitted variables. Instrumental variable methodology is a general approach to improve the causal interpretation of coefficients in a statistical model (Angrist & Krueger, 2001; Angrist, Imbens, & Rubin, 1996; Bound, Jaeger, & Baker, 1995; Hanushek & Jackson, 1977, p. 234–9; Shadish, Cook, & Campbell, 2002; Stolzenberg & Relles, 1990). In mediation, an estimate of the causal relation between M and Y is obtained by using X as an instrumental variable. An instrumental variable makes a relationship more like an experimental relationship, and ideally it is equivalent to a randomized experiment. For the mediation model, the idea is to use an instrument X for the prediction of M and then use the predicted values of M to predict Y. The statistical significance of the coefficient relating predicted value of M to Y is the causal test of the b path for the mediated effect. There are several assumptions required for the b path to have a causal interpretation, two of which are described here (*see* MacKinnon, 2008; Shadish, Cook, & Campbell, 2002). First, it is assumed that the randomizing qualities of the instrumental variable X lead to randomization of the mediator M. The stronger the relation of X to M, the better the instrument, with the ideal instrument having a correlation of 1 for X with M. When the correlation is 1, X and M can be considered the same variable and so the randomization of X allows M to reflect randomization. Because M can be considered randomized, the influence of any omitted variable on the relation of M to Y is removed. Another assumption is that M completely mediates the effect of X on Y. In the instrumental variables literature, this assumption is called the exclusion restriction and it means that there is complete mediation. However, complete mediation is rare and may be unrealistic in many research contexts. If complete mediation is not plausible for the entire sample, then it may be possible to design a research study to identify subgroups where the assumption of complete mediation is reasonable. The applicability of the instrumental variable methods is likely limited because of the requirement of a strong relation between the instrument and M as well as the need for complete mediation. More applications of instrumental variable mediation methods are needed to assess the benefits of this approach.

Principal Stratification

Several promising methods to strengthen causal inference in mediation analysis are based on the classification of different possible response patterns for how X affects M and M affects Y. This approach specifies subsets of persons based on how the relation between M and Y could change in response to treatment X. For example, in a treatment/control study there are four different types of hypothetical responses (Jo, 2008): (1) never-improvers, whose mediator would not improve if they were in either treatment group; (2) forward-improvers, whose mediator would improve only if they received the treatment; (3) backward-improvers, whose mediator would improve only if they were in the control group; and (4) always-improvers, whose mediator would improve if they were in either treatment group. These four types of persons are determined outside the experiment so their classification is independent of their experimental assignment. Typically, it is assumed that there are no backward-improvers because it is often difficult to conceive of situations where the mediator would improve for participants in the control condition. Once the classifications are made, the mediated effect is estimated within and between these stratifications (Angrist, Imbens, & Rubin, 1996; Frangakis & Rubin, 2002; Jo, 2008). In these models, covariates are often used to provide information about the stratifications to identify parameters in the model (Jo, Stuart, MacKinnon, & Vinokur, 2010; Stuart, Perry, Le, & Ialongo, 2008). Related approaches use the actual and counterfactual data for each participant and estimate model parameters in the context of the additional counterfactual condition in which the participants actually did not partake (Robins, 1989; Robins & Greenland, 1992; Witteman et al., 1998). In other words, these counterfactual models consider how control participants would behave if, in fact, they were in the treatment condition and how treatment participants would behave if, in fact, they were in the control condition. Like other causal inference methods, the application of these methods to many real data sets is critical to assess their usefulness for uncovering mediation.

Experimental Designs

Experimental studies can be used to bolster evidence for a mediation relation. A cumulative experimental approach is the most widely used method to identify and validate mediating variables in psychology and other substantive areas. As recognized by researchers, one or a few experiments are not enough to convincingly demonstrate a true mediation relation. The literature suggests research designs to directly test the mediation relation (MacKinnon, 2008; Mark, 1986; West, Aiken, & Todd, 1993; West & Aiken, 1997). The typical research design to assess mediation has a randomized intervention, and the mediator and outcome are measured in each group. An assumption of this design is that the relation of M to Y represents a real causal relation so that directly changing M will lead to a change in Y.

There are other possible experimental designs to assess mediation (MacKinnon, 2008). One such research design is called the blockage design, where a manipulation is used to block the mediation process (Robins & Greenland, 1992). If the intervention to block the mediation process does remove the mediation relation, then this provides additional evidence for the mediation process. As a hypothetical example of the blockage design, consider a study to investigate whether an intervention reduces drug use by changing social norms among friends. Intervention participants may be randomly assigned to a treatment condition where contact among friends was eliminated or to a control condition that allowed regular contact among friends. If social norms among friends is a mediator of the drug prevention program, then reduced drug use should be observed in the control group but not in the treatment group, where norm change was not possible because of the lack of contact among friends.

An enhancement design is similar to a blockage design, with the exception that mediation effects are enhanced, rather than eliminated, in the treatment group (for examples, *see* Maxwell, Bashook, & Sandlow, 1986, and Klesges, Vasey, & Glasgow, 1986). For example, in the drug use intervention study described above, social contact among friends would be enhanced in the treatment group. If social norms among friends are a mediator of the drug use intervention, then the treatment group should show a greater reduction in drug use than the control group. Another type of experimental design called double randomization uses one randomized study to evaluate the X–M relation and a second randomized study to evaluate the M–Y relation adjusting for X (MacKinnon, 2008; MacKinnon & Pirlott, 2010; Spencer et al., 2005). Other experimental designs focus on testing the consistency and specificity of mediation relations across different contexts, subgroups, and measures of the mediating and outcome variables (MacKinnon & Pirlott, 2010).

Longitudinal Mediation

The mediation model is a longitudinal model where X precedes M and M precedes Y. Previous research has demonstrated the limitations of using cross-sectional data to investigate longitudinal mediation processes (Cole & Maxwell, 2003; Maxwell & Cole, 2007). Repeated measures are a feature of well-designed studies because they enhance power to detect program effects and allow for the measure of change in response to a treatment (Cohen, 1988; Singer & Willett, 2003). The importance of temporal precedence in the investigation of mediation has been emphasized (Gollob & Reichardt, 1991; Judd & Kenny, 1981; Kraemer et al., 2002; MacKinnon, 1994) and methods for assessing longitudinal mediation have been described (Cheong et al., 2003; Cole & Maxwell, 2003; MacKinnon, 2008; Maxwell & Cole, 2007). As a result, the evaluation of longitudinal mediation models is an important step in advancing mediation methods. Several choices for longitudinal mediation models have been described, including autoregressive, latent growth curve, and latent change score models. Latent growth curve models are a common choice for analyzing repeated measures data and have been applied to longitudinal mediation models (Cheong, 2002; Cheong et al., 2003). Although autoregressive mediation models (e.g., Cole & Maxwell, 2003) and latent change score mediation models (MacKinnon, 2008) have been outlined, much work remains to be done to evaluate how these models perform in the analysis of real data. It is likely that these different modeling approaches are suited to different research situations.

Two Wave Models

The simplest type of longitudinal study has two waves of data, such as measurement at baseline and follow-up. These two-wave models differ in several ways from the single-wave X, M, and Y models that have been the most thoroughly studied to date. The most common two-wave methods are: (1) analysis of covariance (ANCOVA), (2) difference score analysis, and (3) residualized change score analysis (Bonate, 2000; Tornqvist, Vartia, & Vartia, 1985; Willett & Sayer, 1994). Analysis of covariance includes baseline measures of the variables as predictors of follow-up measures in each mediation equation to adjust for baseline differences. Difference scores and residualized change scores are options for single-mediator models because the two waves of measurement can be reduced to a single change score. In a difference score model, X is the change in X from baseline to follow-up, M is the change in M from baseline to follow-up, and Y is the change in Y from baseline to follow-up. Using the difference score as a dependent variable has been controversial because measurement error may be too high (Burr & Nesselroade, 1990; Cronbach & Furby, 1970; Rogosa, 1988). The residualized change score is often used as an alternative to the difference score and ANCOVA methods, at least in part because it adjusts for baseline differences and avoids some of the problems with the unreliability of difference scores (Lord, 1963). Often the conclusions based on the residualized change score are indistinguishable from the conclusions based on ANCOVA because both approaches adjust for baseline measurement.

Three (or More)-Wave Models

One limitation of the analysis of two waves of data is that the relation of the mediator to the dependent variable is still a cross-sectional relation. If three waves of data are collected, then the longitudinal relation of X and M and the longitudinal relation of M to Y can be examined. With many repeated measures, more accurate modeling of growth over time can be assessed prior to and after the intervention. If more than two waves of data collection are available, there are several longitudinal mediation models that can be applied, including the autoregressive model (Cole & Maxwell, 2003; Maxwell & Cole, 2007), the latent growth curve model (Cheong, MacKinnon, & Khoo, 2001, 2003), and the latent change score model (McArdle, 2001; *see* MacKinnon, 2008). There are additional complexities for three and more wave models, including assumptions regarding the stability of scores over time, stationarity (the same relations among variables across time), equilibrium (the relations among variables are stable enough to be estimated), and timing (the timing of how variables change across time is correct and measured at the right time to detect effects). Space limitations preclude further discussion of these topics, but interested readers can consult MacKinnon (2008). Although longitudinal mediation models are complex, they provide potentially more accurate information regarding the relations among variables (Cohen, 2008; Cole & Maxwell, 2003; MacKinnon, 1994, 2008; Jöreskog, 1979).

Autoregressive Models. The autoregressive mediation model is an extension of the model described in Gollob and Reichardt (1991) and elaborated on by Cole and Maxwell (2003). It specifies the dependency between adjacent longitudinal relations, in addition to the relations consistent with longitudinal mediation. There are several important aspects of this model. First, relations one lag (i.e., one measurement wave) apart are specified. With three waves of data, it is possible to consider effects that are two waves apart as lag-two effects, but these effects are usually not included in the mediation model. Second, examining the relations within a variable over time assesses the stability of that measure. Third, only longitudinal relations consistent with longitudinal mediation are specified among the variables (i.e., X1 is related to M2, and M2 is related to Y3). Fourth, the covariances among variables at the first wave are included (e.g., the covariance between X1 and Y1), as are the covariances among the residuals of X, M, and Y at each later wave (e.g., the covariance between the X2 residuals and the Y2 residuals). This reflects the unknown causal order of these measures within the same wave.

Other forms of the autoregressive model include contemporaneous mediation relations among X, M, and Y (e.g., X1–M1 and M1–Y1), as well as the longitudinal mediation effects described above. Another option for an autoregressive longitudinal mediation model allows for cross-lagged relations among variables (e.g., X1–M2 and M1–X2) so that the direction of relations among X, M, and Y are all free to vary. This model violates the assumed temporal precedence of X to M to Y, as specified by the mediation model, because paths in the opposite direction are also estimated (i.e., M–X and Y–M). However, this model could be used to assess the possibility of cross-lagged relations among variables.

Latent Growth Curve Models. The latent growth curve approach models longitudinal data in a different manner than the autoregressive approach (Duncan, Duncan, Strycker, Li, & Alpert, 2006). In a simple growth model, there are two growth parameters: (1) the random intercept factor, representing the initial status of the growth trajectory at Time 1, and (2) the random slope factor, representing the linear growth rate per time unit. The mediation process includes the relations between the slope factor of the independent variable, the slope factor of the mediating variable, and the slope factor of the outcome variable (Cheong, MacKinnon, & Pentz, 2002; MacKinnon, 2008; Hertzog, Lindenberger, Ghisletta, & Oertzen, 2006). Linear, quadratic, cubic, and higher order trends can be estimated to reflect the growth over time. A two-stage piecewise parallel process latent growth model provides a way to evaluate the effect of earlier change in the growth of the mediator on later change in the growth of the outcome variable.

The latent growth curve framework (Cheong, MacKinnon, & Khoo, 2001, 2003) models the growth curves of the mediator and the outcome as distinct parallel processes influenced by a program. This method has been applied to the evaluation of several prevention studies (e.g., Cheong, MacKinnon, & Khoo, 2001, 2003; Cheong, MacKinnon, & Pentz, 2002; Cheong & MacKinnon, 2008). Cheong (2011) found that testing mediation in the latent growth curve framework requires large sample sizes to obtain accurate estimates and adequate statistical power. For example, a sample size of 500 is needed when the size of the true mediated effect is moderate (i.e., proportion mediated of 0.30). The accuracy of estimates and the statistical power for testing mediation in the latent growth curve framework were also influenced by the fit of the hypothesized trajectory shape to the data; better fit in the growth trajectory portion of the model led to improved estimates and statistical power. More work is needed on the performance of different latent growth curve approaches to assess mediation.

Latent Change Score Models. The latent change score model is related to the latent growth model. A latent change score model allows a researcher to examine the change in a variable between pairs of measurement waves (Ferrer & McArdle, 2003, 2010; McArdle, 2001; McArdle & Hamagami, 2001; McArdle & Nesselroade, 2003). In this model, fixed parameters and modeled latent variables are used to specify latent change scores. By specifying latent change, the model represents dynamic change in a variable (i.e., the acceleration or deceleration of the change in that variable between measurement waves). This model may be especially useful in situations where it is expected that the predictors of change differ based on the wave of measurement. One common example of these effects may occur in experimental research where the manipulation affects change early in the process, but these effects may not be present later in the process (i.e., at later waves). That is, the intervention affects change in a dependent variable between wave 1 and wave 2, but the intervention does not affect change in the dependent variable between later waves (e.g., between waves 3 and 4). The original latent change score model is constrained to represent change between two waves but it is possible to specify models that represent the acceleration/deceleration of the change between waves (i.e., second derivatives; Malone, Lansford, Castellino, Berlin, & Dodge, 2004) and models that represent moving averages. Another version of the latent change score model includes paths relating the latent differences in one variable to the latent differences in another variable. There are many potential mediation effects in this type of model corresponding to mediation effects for concurrent change and mediation effects for longitudinal change, including correlated change between waves and relations among longitudinal change in X, M, and Y.

Person-Centered Approaches

Person-oriented approaches, such as trajectory classes (Muthén & Muthén, 2001), staged response across trials (Collins et al., 1998) and configural frequency analysis (von Eye et al., 2009) represent new ways to understand mediation processes consistent with the goal of examining both individual-level processes and group-level processes. Rather than examining whether relations between variables (e.g., X–M and M–Y adjusted for X) are consistent with mediation, this approach considers whether patterns of data for individual persons are consistent with mediation. A significance test for the person-centered approach has recently been proposed (Fairchild & MacKinnon, 2005; MacKinnon, 2008). These person-centered methods are a welcome addition to traditional variable-centered mediation analysis (von Eye, Mun, & Mair, 2009).

Summary and Future Directions

Many substantive theories specify mediating mechanisms by which an antecedent variable is related to an outcome variable. Although these processes are important for observational studies, it is their application in intervention research, which has led to the recent growth in statistical and methodological research on mediation methods. Action theory, the relation of intervention exposure to the mediator, and conceptual theory, the relation of the mediator to the outcome, are two important theoretical components of the mediation approach in intervention design and evaluation. Tests of action and conceptual theory reveal if and how an intervention program works to change outcomes. Over the last 30 years, new approaches to statistical mediation analysis have been developed. The purpose of this chapter was to describe developments in four major areas: (1) significance testing and confidence interval estimation of the mediated effect, (2) mediation analysis in groups, (3) assumptions of and approaches to causal inference for mediation, and (4) longitudinal mediation models.

The mediated effect is estimated using the difference in coefficients, $c-c'$, method or the product of coefficients, *ab*, method. The single mediator model can easily be expanded to include covariates or additional mediators. Significant tests and confidence intervals based on normal theory result in reduced statistical power and inflated Type I error rates, primarily because the distribution of the mediated effect is not normal. More accurate statistical tests and confidence intervals can be obtained with methods that are based on the distribution of the product or that incorporate the non-normal distribution of the mediated effect (e.g., bootstrapping methods). Bayesian methods also incorporate the non-normal distribution of the mediated effect and incorporate prior information, which may be particularly useful when estimating mediation models in small samples. For these and other reasons, Bayesian methods may be especially useful for mediation analysis and merit further research attention. Effect sizes measures in mediation analysis gage the practical meaning of effects in complement to significance testing. Correlation measures of effect size are useful for individual paths in the mediation model. The proportion mediated, ratio of the indirect effect to the direct effect (ab/c'), and mediated effect divided by the standard deviation of Y are promising effect size measures for the entire mediated effect. Methods for mediation analysis of count and categorical outcomes are available, along with methods for non-normal variables in the mediation model. One issue in testing mediation effects is that relatively large sample sizes are needed. Power can be improved through study design, measurement, and statistical analysis.

Mediation effects may differ across groups such as age or gender (i.e., moderator variables). Mediation describes how an antecedent variable changes an outcome, and moderation describes for whom the antecedent variable changes an outcome. The general moderation and mediation model allows for tests of moderation of a mediated effect (i.e., the mediated effects differs across subgroups of participants) and mediation of a moderated effect (i.e., the mediator is intermediate in the causal sequence from an interaction effect to the outcome). When there are more than a few subgroups in the sample (e.g., schools), using a categorical moderator is not ideal. Rather, the multilevel modeling framework can be applied to mediation models. These multilevel mediation models are interesting because the independent variable, the mediator, and the outcome can be specified at the individual level or the group (i.e., cluster) level, or at both levels. The development of models that combine mediation and subgroup analyses is ongoing.

Causal inference is probably the most rapidly growing area in mediation methodology. When X is a randomly assigned variable, the relations of X to M and X to Y can be interpreted as causal effects. The M–Y relation, however, is not directly randomized so causal interpretation is suspect. The threat of omitted variables to the M–Y relation is a violation of the Sequential Ignorability B assumption, but sensitivity analysis and instrumental variable analysis can address this violation. Applying a principal stratification approach to the mediation model is another way to strengthen causal inference in mediation. Cumulative experimental studies can validate mediating variables. The rapid developments in causal inference for mediation models will likely continue. Ideally, a clear set of conceptual and statistical approaches for causal inference in mediation will emerge.

The mediation model is a longitudinal model where X precedes M and M precedes Y. Temporal precedence is central to a mediation model. Although cross-sectional data have been used to investigate mediation processes, the utility of these models is limited. When there are only two waves of data, the most common methods used to examine mediation effects are ANCOVA, difference score, and residualized change score. For three or more waves of data, several types of longitudinal mediation models have recently been developed, including autoregressive, latent growth curve, and latent change score models. Person-centered approaches to mediation have also begun to emerge in the literature. The next step in research will be to evaluate the performance of these proposed models with real data. It is likely that different modeling approaches are suited to different research situations.

Mediation analyses have the potential to test the theoretical questions commonly posed in many fields. In treatment and prevention research, mediation analysis can identify critical intervention components, thereby reducing intervention costs and increasing scientific understanding of behavior. Mediation analysis addresses the fundamental nature of theories that explain processes by which one variable affects another variable. This chapter demonstrates the rapid growth in methods to identify mediating variables. There is no reason to believe that the demand for accurate mediation analysis or the development of these methods will decline.

Acknowledgments

This article was supported in part by National Institute on Drug Abuse grant DA09757. We thank several members of the Research in Prevention Laboratory for comments on the manuscript.

Author Note

1. David P. MacKinnon, Ph.D., Professor Department of Psychology, Arizona State University, 950 S. McAllister, Room 237, Tempe, AZ 85287–1104, E-mail: David.MacKinnon@asu.edu

2. Yasemin Kisbu-Sakarya, M. A., Graduate Research Assistant, Department of Psychology, Arizona State University, 950 S. McAllister, Room 237, Tempe, AZ 85287–1104, E-mail: ykisbu@asu.edu

3. Amanda Gottschall, M.A., Statistical Analyst, Parenting & Family Research Center, University of South Carolina, 1233 Washington St., 2nd Floor Columbia, SC 29208, Office: 803–978–7412, Fax: 803–978–7410, E-mail: Amanda.Gottschall@sc.edu. The chapter was supported in part by National Institute on Drug Abuse grant DA009757.

References

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

Alf, E. F., & Abrahams, N. M. (1975). Use of extreme groups in assessing relationships. *Psychometrika, 40,* 563–572.

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

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

Angrist, J. D., & Krueger, A. B. (2001). Instrumental variables and the search for identification: From supply and demand to natural experiments. *Journal of Economic Perspectives, 15*(4), 69–85.

Arbuckle, J. L. (1997). *AMOS user's guide version 3.6.* Chicago, IL: SmallWaters Corp.

Baranowski, T., Anderson, C., & Carmack, C. (1998). Mediating variable framework in physical activity interventions. How are we doing? How might we do better? *American Journal of Preventive Medicine 15,* 266–297.

Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology, 51,* 1173–1182.

Bauer, D. J., Preacher, K. J., & Gil, K. M. (2006). Conceptualizing and testing random indirect effects and moderated mediation in multilevel models: New procedures and recommendations. *Psychological Methods, 11,* 142–163.

Begg, C. B., & Leung, D. H. Y. (2000). On the use of surrogate end points in randomized trials. *Journal of the Royal Statistical Society, 163,* 15–28.

Bennett, J. A. (2000). Mediator and moderator variables in nursing research: Conceptual and statistical differences. *Research in Nursing & Health, 23,* 415–420.

Bentler, P. M. (1995). *Theory and implementation of EQS: A structural relations program.* Los Angeles, CA: BMDP Statistical Software.

Blalock, H. M. (1969). *Theory construction: From verbal to mathematical formulations.* Englewood Cliffs, NJ: Prentice Hall.

Blalock, H.M. (Ed.) (1971). *Causal models in the social sciences.* Chicago, IL: Aldine-Atherton.

Bollen, K.A. (1987). Total, direct, and indirect effects in structural equation models. In Clifford Clogg (Ed.), *Sociological Methodology* (pp. 37–69). Washington, DC: American Sociological Association.

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

Bollen, K. A., & Stine, R.A. (1993). Bootstrapping goodness-of-fit measures in structural equation models. In K.A. Bollen & J.S. Long (Eds.), *Testing structural equation models* (pp. 111–135). Newbury Park, CA: Sage.

Bonate, P. L. (2000). *Analysis of pretest-posttest designs.* Boca Raton, FL: Chapman & Hall.

Bound, J., Jaeger, D. A., & Baker, R. M. (1995). Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak. *Journal of the American Statistical Association, 90,* 443–450.

Burr, J. A., & Nesselroade, J. R. (1990). Change measurement. In A. von Eye (Ed.), *Statistical methods in longitudinal research, vol. I* (pp. 3–34). New York: Academic Press.

Chen, H. -T. (1990). *Theory-driven evaluations.* Newbury Park, CA: Sage.

Cheong, J. (August, 2002). *Investigation of a method to evaluate mediating mechanisms to reduce adolescent DUI in a school-based prevention program.* Unpublished doctoral dissertation, Arizona State University.

Cheong, J. (2010). Accuracy of estimates and statistical power for testing mediation in latent growth modeling. *Structural Equation Modeling, 18,* 195–211.

Cheong, J. & MacKinnon, D. P. (May, 2008). Accuracy of estimates and statistical power for testing mediation in latent growth modeling. Paper presented at the annual convention of the Association for Psychological Science, Chicago, IL.

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

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

Cheong, J., MacKinnon, D. P., & Pentz, M. A. (June, 2002). *Evaluation of the mediating mechanisms of a drug prevention program to reduce adolescent DUI.* Poster presented at the annual meeting of the Research Society on Alcoholism, San Francisco, CA.

Clogg, C. C., Petkova, E., & Shihadeh, E. S. (1992). Statistical methods for analyzing collapsibility in regression models. *Journal of Educational Statistics, 17,* 51–74.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences.* Hillsdale, NJ: Erlbaum.

Cohen, P. (2008). Applied data analytic techniques for turning points research. New York: Taylor & Francis.

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

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

Coxe, S., & MacKinnon, D. P. (2010). Mediation Analysis of Poisson distributed count outcomes. *Multivariate Behavioral Research, 45,* 1022.

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

Cronbach, L. J., & Furby, L. (1970). How should we measure "change" or should we? *Psychological Bulletin, 74,* 68–80.

Cumming, G., & Finch, S. (2001). A primer on the understanding, use, and calculation of confidence intervals that are based on central and noncentral distributions. *Educational and Psychological Measurement, 61,* 532–574.

Dagne, G. A., Brown, C. H., & Howe, G. W. (2007). Hierarchical modeling of sequential behavioral data: Examining complex association patterns in mediation models. *Psychological Methods, 12,* 298–316.

Davis, J. A. (1985). *The logic of causal order.* In: Sage university paper series on quantitative applications in the social sciences. Beverly Hills, CA: Sage Publications.

Donaldson, S. I. (2001). Mediator and moderator analysis in program development. In S. Sussman (Ed.), *Handbook of program development for health behavior research and practice* (pp. 470–496). Thousand Oaks, CA: Sage.

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

Duncan, O.D. (1975). *Introduction to structural equation models.* New York: Academic Press.

Duncan, T. E., Duncan, S. C., Strycker, L. A., Li, F., & Alpert, A. (2006). *An introduction to latent variable growth curve modeling: Concepts, issues, and applications.* Mahwah, NJ: Erlbaum.

Edwards, J. R., & Lambert, L. S. (2007). Methods for integrating moderation and mediation: A general analytical framework using moderated path analysis. *Psychological Methods, 12,* 1–22.

Elliott, M. R., Raghunathan,T. E., & Li, Y. (2010). Bayesian inference for causal mediation effects using principal stratification with dichotomous mediators and outcomes. *Biostatistics, 11,* 353–372.

Enders, C. (2010). *Applied missing data analysis.* Guilford: New York.

Fairchild, A.J. (2008). A comparison of frameworks for the joint analysis of mediation and moderation effects. Unpublished dissertation.

Fairchild, A. J., & MacKinnon, D. P. (2005). Developing significance tests for a person-centered approach to mediation. Poster presented at the 113th annual convention of the American Psychological Association, Washington, D.C.

Fairchild, A. J., & MacKinnon, D. P. (2009). A General Model for Testing Mediation and Moderation Effects. *Prevention Science, 10,* 87–99.

Fairchild, A. J., MacKinnon, D. P., Taborga, M., & Taylor, A. B. (2009). R^2 Effect-size Measures for Mediation Analysis. *Behavioral Research Methods. 41,* 486–498.

Ferrer, E. & McArdle, J. J. (2003). Alternative structural models for multivariate longitudinal data analysis. *Structural Equation Modeling, 10,* 493–524.

Ferrer, E. & McArdle, J. J. (2010). Longitudinal modeling of developmental changes in psychological research. *Current Directions in Psychological Science, 19,* 149–154.

Finch, J. F., West, S. G., & MacKinnon, D. P. (1997). Effects of sample size and nonnormality on the estimation of mediated effects in latent variable models. *Structural Equation Modeling, 4,* 87–107.

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

Frank, K. A. (2000). Impact of a confounding variable on a regression coefficient. *Sociological Methods and Research, 29,* 147–194.

Freedman, L. S. (2001). Confidence intervals and statistical power of the 'validation' ratio for surrogate or intermediate endpoints. *Journal of Statistical Planning and Inference, 96,* 143–153.

Freedman, L. S., & Schatzkin, A. (1992). Sample size for studying intermediate end-points within intervention trials or observational studies. *American Journal of Epidemiology, 136,* 1148–1159.

Fritz, M. S., & MacKinnon, D. P. (2007). Required sample size to detect the mediated effect. *Psychological Science, 18*(3), 233–239.

Fritz, M. S., Taylor, A., & MacKinnon, D. P. (2010). *When does the bias-corrected bootstrap test of mediation produce elevated Type I error rates and why?* Poster presented at the annual meeting of the Society for Prevention Research, Denver, CO.

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

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

Goodman, L. A. (1962). The variance of the product of K-random variables. *Journal of the American Statistical Association, 57,* 54–60.

Graham, J. W., Hofer, S. M., Donaldson, S. I., MacKinnon, D. P., & Schafer, J. L. (1997). Analysis with missing data in prevention research. In K. Bryant, M. Windle, & S. G. West (Eds.), *The science of prevention: Methodological advances from alcohol and substance abuse research* (pp. 325–365). Washington, DC: American Psychological Association.

Graham, J. W., Hofer, S. M., & MacKinnon, D. P. (1996). Maximizing the usefulness of data obtained with planned missing value patterns: an application of maximum likelihood procedures. *Multivariate Behavioral Research, 31*(2), 197–218.

Graham, J. W., Hofer, S. M., & Piccinin, A. M. (1994). Analysis with missing data in drug prevention research. In Collins, L. & Seitz, L. (Eds.), *National Institute on Drug Abuse Research Monograph Series,* Vol. 142 (pp. 13–62). Washington, DC: National Institute on Drug Abuse.

Graham, J. W., Taylor, B. J., & Cumsille, P. E. (2001). Planned missing data designs in the analysis of change. In L.M. Collins & A.G. Sayer (Eds.), *New Methods for the Analysis of Change* (pp. 335–353). Washington, DC: American Psychological Association.

Hall, P. (1992). On the removal of skewness by transformation. *Journal of the Royal Statistical Society, Series B, 54,* 221–228.

Hanushek, E. A., & Jackson, J. E. (1977). *Statistical methods for social scientists.* New York: Academic Press.

Harlow, L. L., Mulaik, S. A., & Steiger, J. H. (Eds.). (1997). *What if there were no significance tests?* Mahwah, NJ: Erlbaum.

Hertzog, C., Lindenberger, U., Ghisletta, P., & Oertzen, T. (2006). On the power of multivariate latent growth curve models to detect correlated change. *Psychological Methods, 11*(3), 244–252.

Hofmann, D. A., & Gavin, M. B. (1998). Centering decisions in hierarchical linear models: Implications for research in organizations. *Journal of Management, 24,* 623–641.

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

Hoyle, R. H. (1999). *Statistical strategies for small sample research.* Thousand Oaks: Sage.

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

Hyman, H. H. (1955). *Survey design and analysis: Principles, cases, and procedures.* Glencoe, IL: Free Press.

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

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

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

James, L. R., Mulaik, S. A., & Brett, J. M. (2006). A tale of two methods. *Organizational Research Methods,* 9, 233–244.

Jo, B. (2008). Causal inference in randomized experiments with mediational processes. *Psychological Methods, 13,* 314–336.

Jo, B., Stuart, E. A., MacKinnon, D. P., & Vinokur, A. D. (2010). The use of propensity scores in mediation. *Multivariate Behavioral Research, 46,* 425–452.

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

Jöreskog, K. G. (1979). Statistical estimation of structural models in longitudinal-developmental investigations. In J. R. Nesselroade & P. B. Baltes (Eds.), *Longitudinal research in the study of behavior and development* (pp. 303–352), New York: Academic.

Jöreskog, K. G., & Sorbom, D. (2001). *LISREL 8: User's reference guide.* Lincolnwood, IL: Scientific Software.

Judd, C. M., & Kenny, D. A. (1981). *Estimating the effects of social interventions.* New York: Cambridge University Press.

Judd, C. M., Kenny, D. A., & McClelland, G. H. (2001). Estimating and testing mediation and moderation in within-subject designs. *Psychological Methods, 6,* 115–134.

Kaufman, J. S., MacLehose, R. F., & Kaufman, S. (2004). A further critique of the analytic strategy of adjusting for covariates to identify biologic mediation. *Epidemiologic Perspectives & Innovations, 1*(4), 1–13.

Keesling, J.W. (June, 1972). *Maximum Likelihood approached to causal analysis.* Unpublished doctoral dissertation, University of Chicago.

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

Kenny, D. A. (2008). Reflections on mediation. *Organizational Research methods, 11,* 353–358.

Kenny, D. A., Kashy, D. A., & Bolger, N. (1998). Data analysis in social psychology. In D. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *Handbook of social psychology, Volume 1* (pp. 233–265). New York: McGraw-Hill.

Klesges, R. C., Vasey, M. M., & Glasgow, R. E. (1986). A worksite smoking modification competition: Potential for public health impact. *American Journal of Public Health, 76,* 198–200.

Kraemer, H. C., Wilson, T., Fairburn, C. G., & Agras, S. (2002). Mediators and moderators of treatment effects in randomized clinical trials. *Archives of General Psychiatry, 59,* 877–883.

Krantz, D. H. (1999). The null hypothesis testing controversy in psychology. *Journal of the American Statistical Association, 44,* 1372–1381.

Kristal, A. R., Glanz, K. Tilley, B. C., & Li, S. H. (2000). Mediating factors in dietary change: Understanding the impact of a worksite nutrition intervention. *Health Education and Behavior, 27,* 112–125.

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

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

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

Ledermann, T. & Macho, S. (2009). Mediation in dyadic data at the level of the dyads: A structural equation modeling approach. *Journal of Family Psychology, 23,* 661–670.

Li, Y., Bienias, J., & Bennett, D. (2007). Confounding in the estimation of mediation effects. *Computational Statistics & Data Analysis, 51,* 3173–3186.

Lin, D. Y., Psaty, B. M., & Kronmal, R. A. (1998). Assessing the sensitivity of regression results to unmeasured confounders in observational studies. *Biometrics, 54,* 948–963.

Lipsey, M. W. (1993). Theory as method: Small theories of treatments. In L. B. Sechrest & H. G. Scott (Eds.), *Understanding causes and generalizing about them* (pp. 5–38). San Francisco, CA: Jossey-Bass.

Little, R. (2006). Calibrated Bayes: A Bayes/frequentist roadmap. *The American Statistician, 60,* 213–223.

Lockwood, C. M. (2000). *Effects of nonnormality on tests of mediation.* Unpublished master's thesis, Arizona State University.

Lockwood, C. M., & MacKinnon, D. P. (1998). Bootstrapping the standard error of the mediated effect. *Proceedings of the 23rd Annual Meeting of SAS Users Group International* (pp. 997–1002). Cary, NC: SAS Institute.

Lord, F. M. (1963). Elementary models for measuring change. In C. W. Harris (Ed.) *Problems in measuring change* (pp. 21–38). Madison, WI: University of Wisconsin Press.

Lynch, K. G., Cary, M., Gallop, R., & ten Have, T. R. (2008). Causal mediation analyses for randomized trials. *Health Services and Outcome Methodology, 8,* 57–76.

MacCallum, R. C., Zhang, S., Preacher, K. J., & Rucker, D. D. (2002). On the practice of dichotomization of quantitative variables. *Psychological Methods, 7,* 19–40.

MacKinnon, D. P. (1994). Analysis of mediating variables in prevention intervention studies. In A. Cazares & L. A. Beatty (Eds.), *Scientific methods for prevention intervention research. NIDA research monograph 139* (DHHS Pub. No. 94-3631, pp. 127–53). Washington, DC: Supt. of Docs., U. S. Government Printing Office.

MacKinnon, D. P. (2000). Contrasts in multiple mediator models. In J. Rose, L. Chassin, C. C. Presson, & S. J. Sherman (Eds.), *Multivariate applications in substance use research: New methods for new questions* (pp.141–160). Mahwah, NJ: Lawrence Erlbaum.

MacKinnon, D. P. (2001). Commentary on Donaldson: Mediator and moderator analysis in program development. In S. Sussman (Ed.), *Handbook of program development for health behavior research and practice* (pp. 497–500). Thousand Oaks, CA: Sage.

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

MacKinnon, D. P., Cheong, J., Goldberg, L., Williams, J., & Moe, E. (May, 2002). *Mediating mechanisms in the second cohort of a program to prevent anabolic steroid use.* Paper presented at the annual meeting of the Society for Prevention Research, Seattle, WA.

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

MacKinnon, D. P., Fairchild, A. J., & Fritz, M. S. (2007). Mediation analysis. *Annual Review of Psychology, 58,* 593–614.

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

MacKinnon, D. P., Goldberg, L., Clarke, G. N., Elliot, D. L., Cheong, J., Lapin, A., Moe, E. L., et al. (2001). Mediating mechanisms in a program to reduce intentions to use anabolic steroids and improve exercise self-efficacy and dietary behavior. *Prevention Science, 2,* 15–28.

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

MacKinnon, D. P., & Lockwood, C. M. (May, 2001). *An approximate randomization test for the mediated effect.* Paper presented at the annual meeting of the Western Psychological Association Convention, Maui, Hawaii.

MacKinnon, D. P., Lockwood, C. M., Brown, C. H., Wang, W. & Hoffman, J. M. (2007). The intermediate endpoint effect in logistic and probit regression. *Clinical Trials, 5,* 499–513.

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

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

MacKinnon, D. P., & Pirlott, A. (2010). The unbearable lightness of b. Paper presented at the Annual Meeting of the Society for Personality and Social Psychology.

MacKinnon, D. P., Taborga, M. P., & Morgan-Lopez, A. A. (2002). Mediation designs for tobacco prevention research. *Drug and Alcohol Dependence, 68,* S69–S83.

MacKinnon, D. P., Taylor, A. B., Yoon, M., & Lockwood, C. M. (2009). A comparison of methods to test mediation and other intervening variable effects in logistic regression. Unpublished Manuscript.

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

MacKinnon, D. P., Weber, M. D., & Pentz, M. A. (1989). How do school-based drug prevention programs work and for whom? *Drugs and Society, 3,* 125–143.

Malone, P. S., Lansford, J. E., Castellino, D. R., Berlin, L. J., & Dodge, K. A. (2004). Divorce and child behavior problems: Applying latent change score models to life event data. *Structural Equation Modeling, 11,* 401–423.

Manly, B. F. J. (1997). *Randomization, bootstrap, and Monte Carlo methods in biology.* London: Chapman & Hall.

Mark, M. M. (1986). Validity typologies and the logic and practice of quasi-experimentation. In W. M. K. Trochim, (Ed.), *Advances in quasi-experimental design and analysis* (pp. 47–66). San Francisco: Jossey-Bass.

Mathiew, J. E., DeShon, R. P., & Bergh, D. D. (2008). Mediational inferences in organizational research. *Organizational Research Methods, 11,* 203–223.

Maxwell, J. A., Bashook, P. G., & Sandlow, L. J. (1986). Combining ethnographic and experimental methods in educational evaluation: A case study. In D. M. Fetterman & M. A. Pittman (Eds.), *Educational evaluation: Ethnography in theory, practice, and politics,* (pp. 121–143). Newbury Park, CA: Sage.

Maxwell, S. E., & Cole, D. A. (2007). Bias in cross-sectional analyses of longitudinal mediation. *Psychological Methods, 12,* 23–44.

McArdle, J. J. (2001). A latent difference score approach to longitudinal dynamic structural analysis. In R. Cudeck, S. du Toit, & D. Sörbom (Eds.) *Structural equation modeling: Present and future. A festschrift in honor of Karl Jöreskog* (pp. 341–380). Lincolnwood, IL: Scientific Software International.

McArdle, J. J., & Hamagami, F. (2001). Latent difference score structural models for linear dynamic analyses with incomplete longitudinal data. In L. M. Collins & A. G. Sayer (Eds.) *New methods for the analysis of change* (pp. 139–175). Washington, DC: APA.

McArdle, J. J., & Nesselroade, J. R. (2003). Growth curve analysis in contemporary psychological research. In J. Schinka & W. Velicer (Eds.), *Comprehensive handbook of psychology, Vol. II: Research methods in psychology* (pp. 447–480). New York: Pergamon Press.

McClelland, G.H., & Judd, C.M. (1993). Statistical difficulties of detecting interactions and moderator effects. *Psychological Bulletin, 114,* 376–390.

McDonald, R. P. (1997). Haldane's lungs: A case study in path analysis. *Multivariate Behavioral Research, 32,* 1–38.

McGuigan, K., & Langholtz, B. (1988). *A note on testing mediation paths using ordinary least-squares regression.* Unpublished note.

Meeker, W. Q., Cornwell, L. W., & Aroian, L. A. (1981). *Selected tables in mathematical statistics, Vol. 7.* Providence, RI: Institute of Mathematical Statistics.

Merrill, R. (1994). *Treatment effect evaluation in nonadditive mediation models.* Arizona State University Dissertation.

Merrill, R., MacKinnon, D. P., & Mayer, L. (2006). *Estimating mediated moderator effects in treatment studies.* Manuscript in preparation.

Morgan-Lopez, A. A. (2002). *Bias-corrected bootstrap estimation of mediated baseline by treatment interaction effects in intervention trials.* Invited colloquium presented at Quantitative Methodology in Prevention Seminar Series, The Methodology Center, The Pennsylvania State University, University Park, PA, November 8, 2002.

Morgan-Lopez, A. A. (2003). A simulation study of the mediated baseline by treatment interaction effect in preventive intervention trials (Doctoral dissertation, Arizona State University, 2003). *Dissertation Abstracts International, 64,* 4673.

Morgan-Lopez, A. A., Castro, F. G., Chassin, L., & MacKinnon, D. P. (2003). A mediated moderation model of cigarette use among Mexican-American youth. *Addictive Behaviors, 28,* 583–589.

Morgan-Lopez, A. A., & MacKinnon, D. P. (May-June, 2001). *A mediated moderation model simulation: Mediational processes that vary as a function of second predictors.* Paper submitted for presentation at the 9th Annual Meeting of the Society for Prevention Research, Washington, DC.

Morgan-Lopez, A. A., & MacKinnon, D. P. (2006). Demonstration and evaluation of a method for assessing mediated moderation. *Behavior Research Methods, 38,* 77–87.

MTA Cooperative Group. (1999). Moderators and mediators of treatment response for children with attention-deficit/Hyperactivity disorder. *Archives of General Psychiatry, 56,* 1088–1096.

Murphy, S. A., Van der Laan, M. J., Robins, J. M. & Conduct Problems Prevention Research Group. (2001). Marginal mean models for dynamic regimes. *Journal of the American Statistical Association, 96,* 1410–1423.

Muthén, L. K., & Muthén, B. O. (2001). *Mplus: The comprehensive modeling program for applied researchers.* Los Angeles, CA: Muthén & Muthén.

Noreen, E. W. (1989). *Computer-intensive methods for testing hypothesis: An introduction.* New York: Wiley & Sons.

Pearl, J. (2009). *Causality: Models, reasoning, and inference.* Cambridge, England: Cambridge University Press.

Pearl, J. (2010a). The foundations of causal inference. *Sociological Methodology, 40,* 75–149.

Pearl, J. (2010b). *The mediation formula: A guide to the assessment of causal pathways in non-linear models.* (Tech. Rep.) Los Angeles, CA: University of California, Los Angeles.

Pillow, D. R., Sandler, I. N., Braver, S. L., Wolchik, S. A., & Gersten, J. C. (1991). Theory-based screening for prevention: Focusing on mediating processes in children of divorce. *American Journal of Community Psychology, 19,* 809–836.

Pitts, S. (1997). *The utility of extreme groups analysis to detect interactions in the presence of correlated predictor variables.* Masters document. Arizona State University.

Pituch, K., Stapleton, L. M., & Kang, J. Y. (2006). A comparison of single sample and bootstrap methods to assess mediation in cluster randomized trials. *Multivariate Behavioral Research, 41,* 367–400.

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

Preacher, K. J., Rucker, D. D., & Hayes, A. F. (2007). Addressing moderated mediation hypotheses: Theory, Methods, and Prescriptions. *Multivariate Behavioral Research, 42,* 185–227.

Preacher, K. J., Rucker, D. D., MacCallum, R. C., & Nicewander, W. A. (2005). Use of the extreme groups approach: a critical reexamination and new recommendations. *Psychological Methods, 10,* 178–192.

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

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

Robins, J.M. (1989). The control of confounding by intermediate variables. *Statistics in Medicine, 8,* 679–701.

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

Robins, J. M., Mark, S. D., & Newey, W. K. (1992). Estimating exposure effects by modeling the expectation of exposure conditional on confounders. *Biometrics, 48,* 479–495.

Rogosa, D. R. (1988). Myths about longitudinal research. In K. W. Schaie, R. T. Campbell & W. M. Meredith & S. C. Rawlings (Eds.), *Methodological issues in aging research.* (pp. 171–209). New York: Springer.

Rosenbaum, P. R. (2002). *Observational studies.* New York : Springer.

Rosenberg, M. (1968). *The logic of survey analysis.* New York : Basic Books.

Rozeboom, W.W. (1956). Mediation variables in scientific theory. *Psychological Review, 63,* 249–264.

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology, 66,* 688–701.

Rubin, D.B. (2004). Direct and indirect causal effects via potential outcomes. *Scandinavian Journal of Statistics, 31,* 161–170.

Sandler, I. N., Wolchik, S. A., MacKinnon, D. P., Ayers, T. S., & Roosa, M. W. (1997). Developing linkages between theory and intervention in stress and coping processes. In S. A. Wolchik & I. N. Sandler (Eds.), *Handbook of children's coping: Linking theory and intervention.* (pp. 3–40) New York: Plenum Press.

Schafer, J. L. (1997). *Analysis of incomplete multivariate data.* London: Chapman & Hall.

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

Shipley, B. (2000). *Cause and correlation in biology.* Cambridge, England: Cambridge University Press.

Sidani, S., & Sechrest, L. (1999). Putting program theory into operation. *American Journal of Evaluation, 20,* 227–238.

Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence.* London: Oxford University Press.

Sobel, M. E. (1982). Asymptotic confidence intervals for indirect effects in structural equation models. In S. Leinhardt (Ed.), *Sociological Methodology* (pp. 290–312). Washington, DC: American Sociological Association.

Sobel, M. E. (1986). Some new results on indirect effects and their standard errors in covariance structure models. In N. Tuma (Ed.), *Sociological Methodology* (pp. 159–186). Washington, DC: American Sociological Association.

Sobel, M. E. (1998a). Causal inference in statistical models of the process of socioeconomic achievement: A case study. *Sociological Methods & Research, 27,* 318–348.

Sobel, M. E. (March, 1998b). *Causality: New ideas on mediated effects.* Paper presented at Arizona State University's Conference on Mediation Models in Prevention Research, Tempe, AZ.

Sobel, M.E. (2007). Identification of causal parameters in randomized studies with mediating variables. *Journal of Educational and Behavioral Statistics, 33,* 230–251.

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

Springer, M. D. (1979). *The algebra of random variables.* New York: Wiley.

Springer, M. D., & Thompson, W. E. (1970). The distribution of products of Beta, Gamma, and Gaussian variables. *Journal of SIAM Applied Mathematics, 18,* 721–737.

Stolzenberg, R. M. (1979). The measurement and decomposition of causal effects in nonlinear and non additive models. In K. F. Schuessler (Ed.), *Sociological Methodology* (pp. 459–488). San Francisco, CA: Jossey-Bass.

Stolzenberg, R. M., & Relles, D. A. (1990). Theory testing in a world of constrained research design. *Sociological Methods & Research, 18*(4), 395–415.

Stuart, E. A., Perry, D. F., Le, H., & Ialongo, N. S. (2008). Estimating intervention effects of prevention programs: Accounting for noncompliance. *Prevention Science, 9,* 288–298.

Taylor, A. B., & MacKinnon, D. P. (2012). Four applications of permutation methods to testing a single-mediator model. *Behavior Research Methods,* xxx–xxx.

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

Tein, J.-Y., & MacKinnon, D. P. (2003). Estimating mediated effects with survival data. H. Yanai, A. O. Rikkyo, K. Shigemasu, Y. Kano, J. J. Meulman (Eds.) *New Developments on Psychometrics* (pp. 405–412). Tokyo, Japan: Springer-Verlag Tokyo.

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

Tornqvist, L., Vartia, P., & Vartia, Y. O. (1985). How should relative changes be measured? *American Statistician, 39*, 43–46.
VanderWeele, T. J. (2008). Simple relations between principal stratification and direct and indirect effects. *Statistics and Probability Letters, 78*, 2957–2962.
VanderWeele, T. J. (2010). Bias formulas for sensitivity analysis for direct and indirect effects. *Epidemiology, 21*, 540–51.
von Eye, A., Mun, E. Y., & Mair, P. (2009). What carries a mediation process? Configuralanalysis of mediation. *Integration of Psychology and Behavior, 43*, 228–247.
Weiss, C. H. (1997). How can theory-based evaluation make greater headway? *Evaluation Review, 21*, 501–524.
West, S. G., & Aiken, L. S. (1997). Toward understanding individual effects in multiple component prevention programs: Design and analysis strategies. In K. Bryant, M. Windle, & S. G. West (Eds.), *The science of prevention: Methodological advances from alcohol and substance abuse research* (pp. 167–209). Washington, DC: American Psychological Association.
West, S. G., Aiken, L. S., & Todd, M. (1993). Probing the effects of individual components in multiple component prevention programs. *American Journal of Community Psychology, 21*, 571–605.
West, S. G., Sandler, I., Baca, L., Pillow, D., & Gersten, J. C. (1991). The use of structural equation modeling in generative research: Toward the design of a preventive intervention for a bereaved children. *American Journal of Community Psychology, 19*, 459–480.
Wiley, D.E. (1973). The identification problem for structural equation models with unmeasured variables. In A.S. Goldberger & O.D. Duncan (Eds.), *Structural equation models in the social sciences* (pp. 69–83). New York: Seminar Press.
Willett, J. B., & Sayer, A. G. (1994). Using covariance structure analysis to detect correlates and predictors of individual change over time. *Psychological Bulletin, 116*, 363–381.
Williams, J., & MacKinnon, D.P. (2002). *Selection bias and the estimation of the mediated effect.* Poster presented at the Western Psychological Association conference, Irvine, CA.
Williams, J., & MacKinnon, D. P. (2008). Resampling and distribution of the product methods for testing indirect effects in complex models. *Structural Equation Modeling, 15*, 23–51.
Winship, C., & Mare, R. D. (1983). Structural equations and path analysis for discrete data. *American Journal of Sociology, 89*, 54–110.
Winship, C., & Morgan, S. L. (1999). The estimation of causal effects from observational data. *Annual Review of Sociology, 25*, 659–707.
Witteman, J. C. M., D'Agostino, R. B., Stijnen, T., Kannel, W. B., Cobb, J. C., de Ridder, M. A. J., Hofman, A., & Robins, J. M. (1998). G-estimation of causal effects: Isolated systolic hypertension and cardiovascular death in the Framingham Heart Study. *Journal of Epidemiology, 148*, 390–401.
Wright, S. (1920). The relative importance of heredity and environment in determining the piebald pattern of guinea-pigs. *Proceedings of the National Academy of Sciences, 6*, 320–332.
Wright, S. (1921). Correlation and causation. *Journal of Agricultural Research, 20*, 557–585.
Yuan, Y., & Mackinnon, D. P. (2009). Bayesian Mediation Analysis. *Psychological Methods, 14*, 301–322.
Zhang, Z., Zyphur, M. J., & Preacher, K. J. (2009). Testing multilevel mediation using hierarchical linear models. *Organizational Research Methods, 12*, 695–719.
Zu, J., & Yuan, K. (2010). Local influence and robufst procedures for mediation analysis. *Multivariate Behavioral Research, 45*, 1–44.

CHAPTER 17

Moderation

Herbert W. Marsh, Kit-Tai Hau, Zhonglin Wen, Benjamin Nagengast, *and* Alexandre J.S. Morin

Abstract

Moderation (or interaction) occurs when the strength or direction of the effect of a predictor variable on an outcome variable varies as a function of the values of another variable, called a moderator. Moderation effects address critical questions, such as under what circumstances, or for what sort of individuals, does an intervention have a stronger or weaker effect? Moderation can have important theoretical, substantive, and policy implications. Especially in psychology with its emphasis on individual differences, many theoretical models explicitly posit interaction effects. Nevertheless, particularly in applied research, even interactions hypothesized on the basis of strong theory and good intuition are typically small, nonsignificant, or not easily replicated. Part of the problem is that applied researchers often do not know how to test interaction effects, as statistical best practice is still evolving and often not followed. Also, tests of interactions frequently lack power so that meaningfully large interaction effects are not statistically significant. In this chapter we provide an intuitive overview to the issues involved, recent developments in how best to test for interactions, and some directions that further research is likely to take.

Key Words: Interaction effect; moderator, moderated multiple regression; mediation; latent interaction; product indicator; structural equation model

Introduction

Moderation and interactions between variables are important concerns in psychology and the social sciences more generally (here we use moderation and interaction interchangeably). In educational psychology, for example, it is often hypothesized that the effect of an instructional technique will interact with characteristics of individual students, an aptitude-treatment interaction (Cronbach & Snow, 1979). For example, a special remediation program developed for slow learners may not be an effective instructional strategy for bright students (i.e., the effect of the special remediation "treatment" is moderated by the ability "aptitude" of the student). Developmental psychologists are frequently interested in how the effects of a given variable are moderated by age in longitudinal or cross-sectional studies (i.e., effects interact with age or developmental status). Developmental psychopathologists may also be interested to know whether a predictor variables is, in fact, a risk factor, predicting the emergence of new symptoms and useful in preventive efforts, or an aggravation factor, mostly useful in curative efforts (i.e. effects interact with the baseline level on the outcome in longitudinal studies; Morin, Janoz, & Larivée, 2009). Social psychologists and sociologists are concerned with how the effects of individual characteristics are moderated by groups in which people interact with others. Organizational psychologists study how the effects of individual employee characteristics interact with workplace characteristics. Personnel psychologists want to know whether a selection test is equally valid at predicting work performance for different