
[All ETDs from UAB](#)

[UAB Theses & Dissertations](#)

2011

An Examination of Statistical Methods for Longitudinal Mediation Modeling

John Paul Bentley
University of Alabama at Birmingham

Follow this and additional works at: <https://digitalcommons.library.uab.edu/etd-collection>



Part of the [Public Health Commons](#)

Recommended Citation

Bentley, John Paul, "An Examination of Statistical Methods for Longitudinal Mediation Modeling" (2011).
All ETDs from UAB. 1145.
<https://digitalcommons.library.uab.edu/etd-collection/1145>

This content has been accepted for inclusion by an authorized administrator of the UAB Digital Commons, and is provided as a free open access item. All inquiries regarding this item or the UAB Digital Commons should be directed to the [UAB Libraries Office of Scholarly Communication](#).

AN EXAMINATION OF STATISTICAL METHODS FOR LONGITUDINAL
MEDIATION MODELING

by

JOHN P. BENTLEY

DAVID L. ROTH, COMMITTEE CHAIR
INMACULADA (CHICHI) B. ABAN
CYNTHIA J. BROWN
LESLIE AIN MCCLURE
GERALD MCGWIN
DAVID T. REDDEN

A DISSERTATION

Submitted to the graduate faculty of The University of Alabama at Birmingham,
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy

BIRMINGHAM, ALABAMA

2011

Copyright by
John P. Bentley
2011

AN EXAMINATION OF STATISTICAL METHODS FOR LONGITUDINAL MEDIATION MODELING

JOHN P. BENTLEY

BIOSTATISTICS

ABSTRACT

The use of mediation modeling is quite prevalent in a number of disciplines to answer questions about how or why one variable exerts its influence on another variable. Although mediation can be assessed in the context of several types of study designs, the use of cross-sectional data and a single-mediator model tend to be the most commonly reported features in empirical tests of mediation. There are several limitations associated with assessing mediation with cross-sectional data, perhaps the most significant is that mediated effect estimates are biased in the case of true longitudinal mediation. For this and several other reasons, there has been a greater emphasis on the development of longitudinal mediation models.

There are several classes of models for evaluating longitudinal mediation with the collection of three or more waves of data. These models are increasingly utilized in the applied literature and methodological research continues to evaluate them, as well as extensions and new approaches. Despite their use in substantive research, the preponderance of mediation hypotheses are still tested with cross-sectional data. Furthermore, consensus on the optimal implementation of longitudinal mediation modeling methods is largely lacking and there are many unanswered questions.

In the first paper of this dissertation we sought to demonstrate the application of one approach to longitudinal mediation modeling, namely the autoregressive model, and build on a set of steps recommended for testing such models. In the other papers we

attempted to address significant methodological questions related to different methods of longitudinal mediation modeling. Paper two explored the effects of unreliability and the failure to account for shared method variance in the autoregressive mediation model. Paper three evaluated the statistical performance of methods used to test mediation in a two-stage piecewise parallel process latent growth curve model and examined the impact of misspecifying the true piecewise model as a single-stage parallel process model. Although mentioned in the literature as a possible method to address a substantial criticism associated with parallel process mediation models (i.e., the inability to delineate temporal ordering), piecewise growth models are rarely utilized in substantive research and lack a full elaboration in the literature.

Keywords: mediation, indirect effects, longitudinal, autoregressive mediation modeling, latent growth curve modeling, parallel process models of mediation

DEDICATION

I would like to dedicate this dissertation to two very special people. To my wife and partner, Sandy, a constant source of love and support, and to my teacher, mentor, and friend, Lon Larson, you are missed but never forgotten.

ACKNOWLEDGMENTS

Although this dissertation may be an individual accomplishment, it would not have been possible without the input, help, and support from many different people. Being “somewhat” of a nontraditional graduate student, in some ways I relied more on others this time, compared to my first time around.

Dr. David Roth has been an outstanding mentor. He knows that one’s growth as a scholar comes not only from formal didactic education, but also by producing new knowledge and that often means trying to “figure things out” on your own. He has been a source of encouragement and support while also pushing and challenging me. I have very much enjoyed learning from him.

I would also like to thank my committee members, Drs. Inmaculada Aban, Cynthia Brown, Leslie McClure, Gerald McGwin, and David Redden. Your time, support, and advice concerning this dissertation and my development in the graduate program are very much appreciated. Thank you also to the UAB Study of Aging researchers who provided access to the data analyzed in paper one of this dissertation. In addition, Drs. Richard Allman and Patricia Sawyer provided helpful suggestions, enhancing the quality of this manuscript.

I am grateful to the faculty, staff, and graduate students in the Department of Biostatistics at the University of Alabama at Birmingham. Over the years, I have had the opportunity to learn from, and work with, some outstanding teachers, researchers, mentors, and students. Fortunately for my current situation, several of those outstanding

individuals are in the UAB Department of Biostatistics. In addition to the four Biostatistics faculty members on my committee, I want to especially recognize Drs. Christopher Coffey and Charles Katholi for their wisdom, support, and guidance. I also want to recognize and thank Ms. Della Daniel for her assistance and kindness.

Thank you to Dr. Patrick Pace from the Center for Pharmaceutical Marketing and Management at the University of Mississippi, who helped me automate the process for running simulations in *Mplus*, saving me countless hours and reminding me about how to use batch files in DOS.

I also want to thank the faculty, staff, and graduate students from the Department of Pharmacy Administration at the University of Mississippi for their understanding, patience, support, and encouragement during the five plus years I spent living a double life in Oxford and Birmingham, especially my department chairs during this period, who were all willing to work with me in the accomplishment of this goal (Drs. Noel Wilkin, Dewey Garner, and Donna West). I especially want to recognize Dr. David McCaffrey, who “volunteered” to assume several of my responsibilities during this time and did them better than I ever could.

I owe a great deal to my parents, Paul and Delphine Bentley, and my siblings, Mark, Lisa, Gina, Sara, and Mary, who provide unconditional love, support, and encouragement. I lost my dad in 2006, but he continues to be a source of inspiration. To my brother Mark, thank you for the many phone conversations while I was driving between Oxford and Birmingham; your humor and words of encouragement will always be remembered and appreciated.

Looking back on this journey, I can recall the quizzical looks from many different people when I tried to explain what I was doing and why I was doing it. Although it made sense to me (I recognized that more formal, basic education and training in statistics would be helpful in my quest to be a life-long learner, as well as a better researcher and teacher), I can understand why some would question the decision to pursue additional graduate education while already engaged in a full-time faculty appointment. One person provided me with unwavering support and encouraged me to continue working toward my goal – my wife, Dr. Sandra Bentley. Not only did she put up with me being gone for several days at a time over a period of several years, she also willingly (I think) participated in numerous discussion about statistics. I am truly thankful and fortunate to have Sandy in my life.

TABLE OF CONTENTS

	<i>Page</i>
ABSTRACT	iii
DEDICATION	v
ACKNOWLEDGMENTS	vi
LIST OF TABLES	xi
LIST OF FIGURES	xiii
INTRODUCTION	1
Overview of Mediation Modeling	3
The General Concept of Mediation.....	3
Mediation in the General Linear Model Framework: Terminology of Mediation Modeling.....	6
Mediation Modeling with Cross-Sectional Data.....	10
Summary of Basic Mediation Modeling.....	12
Longitudinal Mediation Modeling.....	15
Autoregressive Mediation Models.....	18
Random Effects Mediation Models	23
Comparison of Longitudinal Mediation Models.....	34
Latent Variables	36
Summary of Research Objectives.....	39
FUNCTIONAL STATUS, LIFE-SPACE MOBILITY, AND QUALITY OF LIFE: A LONGITUDINAL MEDIATION ANALYSIS	41
EFFECTS OF RANDOM MEASUREMENT ERROR AND FAILURE TO ACCOUNT FOR SHARED METHOD VARIANCE IN TESTING AUTOREGRESSIVE LONGITUDINAL MEDIATION MODELS	73
A SIMULATION STUDY OF TWO-STAGE PIECEWISE LATENT GROWTH CURVE MODELS FOR TESTING LONGITUDINAL MEDIATION	131

SUMMARY, CONCLUSIONS, AND FUTURE RESEARCH.....	185
Concluding Comments on Paper One.....	186
Concluding Comments on Paper Two	187
Concluding Comments on Paper Three	188
Future Research Directions.....	190
Implications for Practice	191
GENERAL LIST OF REFERENCES	193
APPENDIX: UAB INSTITUTIONAL REVIEW BOARD (IRB) APPROVAL FORMS FOR FIRST PREPRINT	199

LIST OF TABLES

<i>Table</i>		<i>Page</i>
FUNCTIONAL STATUS, LIFE-SPACE MOBILITY, AND QUALITY OF LIFE: A LONGITUDINAL MEDIATION ANALYSIS		
1	Descriptive statistics	65
2	Model fit indices and unstandardized path coefficients for cross-sectional mediation models	66
3	Model fit indices and unstandardized path coefficients for longitudinal mediation models	67
EFFECTS OF RANDOM MEASUREMENT ERROR AND FAILURE TO ACCOUNT FOR SHARED METHOD VARIANCE IN TESTING AUTOREGRESSIVE LONGITUDINAL MEDIATION MODELS		
1	Mean relative bias of the mediated effect estimates ($a*b$) across simulation conditions	113
2	Mean relative bias of the $M_1 \rightarrow M_2$ path estimates (path m) across simulation conditions	114
3	Mean relative bias of the $X_1 \rightarrow M_2$ path estimates (path a) across simulation conditions	115
4	Mean relative bias of the $M_2 \rightarrow Y_3$ path estimates (path b) across simulation conditions	116
5	Mean bias of the $X_1 \rightarrow Y_2$ path estimates (path c_1') (i.e., first direct effect) across simulation conditions	117
6	Mean bias of the $X_1 \rightarrow Y_3$ path estimates (path c_2') (i.e., second direct effect) across simulation conditions	118
7	Mean relative bias of standard error of the mediated effect ($a*b$) across simulation conditions	119

8	Estimated Type I error rates for the mediated effect across simulation conditions	120
9	Empirical power levels across simulation conditions for the test of the mediated effect.....	121
10	Estimated Type I error rates for the direct effects across simulation conditions levels.....	122

A SIMULATION STUDY OF TWO-STAGE PIECEWISE LATENT GROWTH CURVE MODELS FOR TESTING LONGITUDINAL MEDIATION

1	Model-implied means for the mediator and the outcome under a few simulation conditions	171
2	Mean fit index values across simulation conditions	172
3	Mean relative bias of the mediated effect estimates ($\hat{\alpha} * \hat{\beta}$) across simulation conditions	173
4	Mean bias of the mediated effect estimates ($\hat{\alpha} * \hat{\beta}$) when mediated effect = 0 across simulation conditions.....	173
5	Mean relative bias of path $\hat{\beta}$ across simulation conditions	174
6	Mean bias of path $\hat{\beta}$ when $\beta = 0$ across simulation conditions	174
7	Mean bias of path $\hat{\tau}'$ across simulation conditions	175
8	Mean relative bias of standard error of the mediated effect ($\hat{\alpha} * \hat{\beta}$) across simulation conditions	176
9	Estimated Type I error rates for the mediated effect across simulation Conditions	177
10	Empirical power levels across simulation conditions for the test of the mediated effect.....	178
11	Estimated Type I error rates and empirical power for the direct effects across simulation conditions levels	179

LIST OF FIGURES

<i>Figure</i>		<i>Page</i>
INTRODUCTION		
1	Path diagram and equations for the basic single-mediator model	6
2	Two-wave autoregressive model	17
3	Basic three-wave autoregressive mediation model.....	19
4	Four-wave autoregressive mediation model	21
5	Basic three-wave latent growth curve model.....	27
6	Three parallel process latent growth curve mediation model with three waves of data for the independent variable (X), the mediator (M), and the dependent variable (Y).....	29
7	Parallel process latent growth curve mediation model with binary X variable (the independent variable) and five waves of data for the mediator (M) and the dependent variable (Y).....	32
8	Two-stage piecewise parallel process latent growth curve mediation model with binary X variable (the independent variable) and five waves of data for the mediator (M) and the dependent variable (Y)	33
FUNCTIONAL STATUS, LIFE-SPACE MOBILITY, AND QUALITY OF LIFE: A LONGITUDINAL MEDIATION ANALYSIS		
1	The Wilson-Cleary health-related quality of life model	68
2	Basic mediation model for the present study	69
3	Autoregressive mediation model framework for the present study	70
4	Completely standardized parameter estimates from the final model (i.e., Model 2) with PCS-12 as the outcome variable including fit-based model enhancements	71

5	Completely standardized parameter estimates from the final model (i.e., Model 2) with MCS-12 as the outcome variable including fit-based model enhancements	72
---	---	----

EFFECTS OF RANDOM MEASUREMENT ERROR AND FAILURE TO ACCOUNT FOR SHARED METHOD VARIANCE IN TESTING AUTOREGRESSIVE LONGITUDINAL MEDIATION MODELS

1	Basic three-wave autoregressive mediation model.....	123
2	Basic three-wave autoregressive mediation model with a latent mediator variable.....	124

A SIMULATION STUDY OF TWO-STAGE PIECEWISE LATENT GROWTH CURVE MODELS FOR TESTING LONGITUDINAL MEDIATION

1	Parallel process latent growth curve mediation model with binary X variable (the independent variable) and five waves of data for the mediator (M) and the dependent variable (Y).....	180
2	Two-stage piecewise parallel process latent growth curve mediation model with binary X variable (the independent variable) and five waves of data for the mediator (M) and the dependent variable (Y)	181

INTRODUCTION

Tests of statistical mediation provide a useful set of techniques for understanding complex relationships among variables. At its most basic level, mediation occurs when an independent variable (X) (also called the putative causal variable) causes an intervening variable or mediator variable (M), which in turn causes the dependent variable (Y). Such models can be used to explore relationships among variables, such as how individuals' attitudes affect behavior through intentions, to evaluate the validity of surrogate endpoints (variables that can be used instead of the ultimate dependent variable), or to assess how or why an intervention produces change on an outcome variable. Understanding the latter may help researchers design more effective treatments and prevention programs by focusing on components of an intervention that change mediator constructs which are ultimately shown to cause changes in an outcome. Such efforts are an attempt to understand the underlying mechanisms of an intervention.

Mediation models can be fit in the context of linear regression, logistic regression and probit regression (see MacKinnon, Lockwood, Brown, Wang, & Hoffman, 2007), path analysis and structural equation modeling (see Bollen, 1987), survival analysis (see Tein & MacKinnon, 2003; Mittelman, Haley, Clay, & Roth, 2006), multilevel modeling (see Krull & MacKinnon, 2001), and other techniques. Decisions regarding the appropriate use of which statistical method depends on a number of factors, including the number and type (i.e., continuous vs. discrete) of the putative causal, mediator, and

outcome variables, the inclusion of latent variables extracted from multiple indicator variables, the presence of correlated data (such as repeated-measures data or students nested within classes), and the nature of the underlying mediation question.

Statistical methods for assessing mediation with continuous variables from a multivariate normal distribution gathered in a cross-sectional manner are well described, usually based on the general linear model (e.g., see MacKinnon, Lockwood, Hoffman, West, & Sheets, 2002). Not surprisingly, the use of cross-sectional data (or longitudinal data treated in a quasi cross-sectional manner, i.e., failing to fully account for the repeated measurement of the mediator and the outcome) is still the norm in most empirical tests of mediation, especially in the field of psychology (Maxwell & Cole, 2007), where mediation modeling is often conducted. The collection of longitudinal data adds potentially improved interpretation, but introduces a number of modeling options and additional considerations (Cole & Maxwell, 2003; Maxwell & Cole, 2007; Collins & Graham, 2002).

There are several approaches available for researchers attempting to assess longitudinal mediational relationships and such methods are being increasingly utilized in the applied literature. Furthermore, methodological research continues to evaluate these models, as well as extensions and new approaches. However, despite their growing use in substantive research, the preponderance of mediation hypotheses are still tested with cross-sectional data. Furthermore, consensus on the optimal implementation of these longitudinal mediation modeling methods is largely lacking and there are still many unanswered questions.

Therefore, the purpose of this dissertation was: 1) to demonstrate the application of one approach to longitudinal mediation modeling, namely the autoregressive mediation model, and build on a set of steps recommended for testing such models; and 2) to address significant methodological questions related to two different classes of longitudinal mediation models, the autoregressive mediation model and the parallel process latent growth curve model of mediation.

The following sections in this chapter review several concepts related to mediation, with an explicit emphasis on longitudinal mediation modeling. While the use of time-to-event variables could be conceptualized to fall under this umbrella, this paper will not discuss mediation modeling in the context of survival analysis. Rather, it will focus on the case where measures of the mediator and outcome variables (and possibly the putative causal variable, although it is possible that that could be a time-invariant variable, such as assignment to experimental condition) are collected on repeated occasions. There are several different models available to researchers to assess mediation in the context of a longitudinal design. The autoregressive model and the random effects model (specifically, the latent growth curve model) form the basis of the methods evaluated in this dissertation and thus will receive the most emphasis.

1.1 Overview of Mediation Modeling

1.1.1 The General Concept of Mediation

In their seminal piece, Baron and Kenny (1986) define a mediator as a variable that accounts for all or part of the relation between a predictor and an outcome. A definition that more clearly delineates the causal nature of mediation relationships comes

from Hoyle and Kenny (1999): “Statistical mediation is present when it can be demonstrated that the causal influence of one variable on another is transmitted through one or more additional variables, referred to as mediators or intervening variables.”

Although the terminology differs (and occasionally the assumptions and statistical tests), the concept of mediation, more broadly labeled the intervening variable effect, is prevalent in many different disciplines. MacKinnon et al. (2002) note that psychology frequently uses the term mediation, sociology uses indirect effect, and epidemiology and the biomedical sciences use surrogate or intermediate endpoint effect. Examples of the latter provided by MacKinnon (2008a) and MacKinnon et al. (2007) include serum-cholesterol as a surrogate endpoint for coronary heart disease and measures of immune system response as a surrogate for death in HIV-infected individuals. In the health literature, some suggest that the criteria for establishing the validity of a surrogate outcome include a finding that the surrogate endpoint explains all of the relation between a treatment and the ultimate dependent variable (in the mediation literature, this is referred to as complete mediation) (Prentice, 1989; Freedman, Graubard, & Schatzkin, 1992). This is not a requirement for establishing mediation in the social sciences, as Baron and Kenny (1986) explicitly allow for the possibility of partial mediation, because rarely in the social sciences can a single variable explain the entire relation between predictor and outcome.

Variables added to models may serve a variety of roles. When a third-variable is added to a model, it can change the interpretation of the relationship between X and Y. These effects are termed third-variable effects and several types are possible. Mediation

as described above is one type of third-variable effect. Confounding and moderation are other commonly examined third-variable effects.

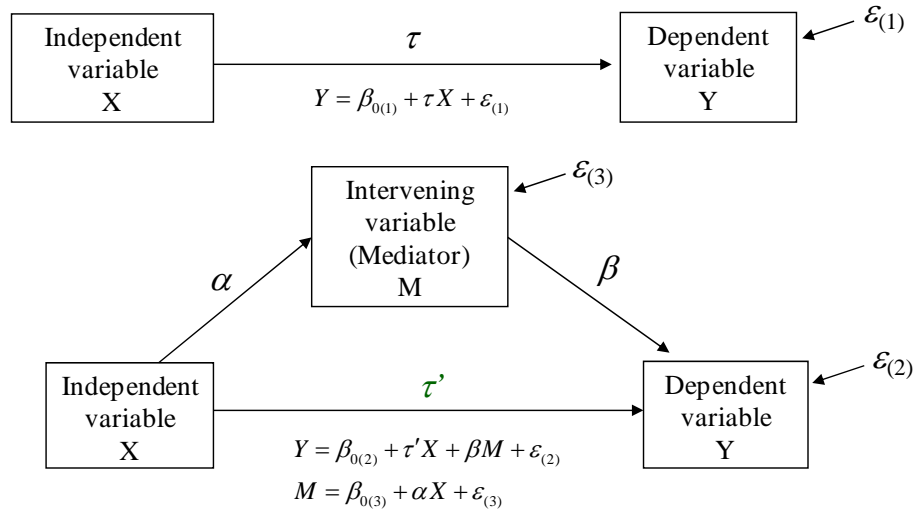
In the context of a cross-sectional design, MacKinnon, Krull, and Lockwood (2000) note that mediation and confounding are mathematically identical concepts and can only be distinguished conceptually. Confounding differs conceptually from mediation in that a confounder is not an intermediate effect in a causal sequence – it is simply related to both X and Y. Mediators transmit the effect of X on Y through the mediator M. Some criteria for defining a confounder explicitly state that the confounder must *not* be an effect of the independent variable – it cannot be part of the causal pathway (Rothman, Greenland, & Lash, 2008). Cole and Maxwell (2003) and Maxwell and Cole (2007) note that distinction among these third-variable effects can only be made in the context of a longitudinal design. The use of cross-sectional data (or longitudinal data treated in a quasi cross-sectional manner, i.e., failing to fully account for the repeated measurement of the mediator and the outcome) is still the norm in most empirical tests of mediation (Maxwell & Cole, 2007). The biases associated with the use of cross-sectional data to assess longitudinal mediation will be discussed in a later section.

The concepts of moderation and mediation are often confused. As mentioned above, mediation is a type of intervening variable model, and a mediation model hypothesizes that X causes M which then causes Y. On the other hand, a moderator variable alters the strength or direction of the relationship between the independent variable and the dependent variable; in essence, the relationship between the X and Y is different at different levels of the moderator (Baron & Kenny, 1986). Moderation

involves the presence of an interaction, whereas mediation implies that the effect of X is transmitted through the mediator.

1.1.2 Mediation in the General Linear Model Framework: Terminology of Mediation Modeling

Before discussing the problems associated with the estimation of longitudinal mediated effects with cross-sectional data and reviewing commonly used longitudinal mediation models, it is useful to review the traditional mediation model in a general linear model framework (i.e., the single mediator, single-level, cross-sectional mediation model with continuous X, M, and Y). Such a review serves to introduce the basic terminology associated with mediation modeling. The necessary regression equations and path models for this model can be found in Figure 1.



Direct effect = τ'

Indirect effect = $\alpha\beta$ ($=\tau - \tau'$ with OLS regression)

Total effect = $\alpha\beta + \tau'$ ($=\tau$ with OLS regression)

Figure 1. Path diagram and equations for the basic single-mediator model.

There are two general approaches to quantify mediation (MacKinnon et al., 2002). The difference in coefficients method estimates the size of the mediated effect (or indirect effect) by taking the difference between the regression coefficients of Y on X before and after adjustment for the mediator ($\tau - \tau'$ in Figure 1). The product of coefficients method estimates the indirect effect as the product of the regression coefficient of M on X and the partial regression coefficient of Y on M adjusted for X ($\alpha\beta$ in Figure 1).¹ The direct effect (τ') represents the effect of X on Y that is not transmitted through M. In the case of a continuous outcome variable with ordinary least squares models and assuming no missing data, MacKinnon, Warsi, and Dwyer (1995) showed that $\tau - \tau'$ is algebraically equivalent to $\alpha\beta$. This same finding is generally not true for other models, such as models with multiple mediators, logistic regression models with a binary mediating or outcome variable, and longitudinal mediation models with repeated observations on mediators or outcomes (e.g., see MacKinnon & Dwyer, 1993; MacKinnon et al., 2007; Roth & MacKinnon, 2012).

Until Sobel (1982) derived an estimate of the standard error of the mediated effect (for product of coefficients method) using the multivariate delta method based on a first-order Taylor series approximation, researchers typically used a series of hypothesis tests to demonstrate the presence of mediation (e.g., the Baron & Kenny (1986) and Judd & Kenny (1981) causal steps methods). Based on a series of simulation studies, MacKinnon et al. (2002) show that such procedures are low in power. In addition, the causal steps methods fail to directly address the mediational hypothesis, rather focusing attention on inferring the presence of indirect effects from patterns in a set of regression

¹Note that in practice, estimates of the population parameters ($\hat{\tau}, \hat{\tau}', \hat{\alpha}, \hat{\beta}$) are used for the calculations.

coefficients (Little, Preacher, Selig, & Card, 2007). Although there are other available standard error formulas, including standard errors for difference in coefficients methods (see MacKinnon et al., 2002 for a review), the Sobel derived standard error is the most commonly used standard error to test for statistical significance of the mediated effect and to construct confidence intervals (using the standard normal distribution):

$$s_{first} = \sqrt{\hat{\alpha}^2 s_{\hat{\beta}}^2 + \hat{\beta}^2 s_{\hat{\alpha}}^2}, \text{ where } s_{\hat{\alpha}}^2 \text{ and } s_{\hat{\beta}}^2 \text{ correspond to the squared standard errors of } \hat{\alpha} \text{ and } \hat{\beta}, \text{ respectively.}$$

To test the null hypothesis of no mediation ($H_0 : \alpha\beta = 0$), the estimate of the mediated effect, $\hat{\alpha}\hat{\beta}$, is divided by the Sobel standard error and this ratio is compared to the standard normal distribution. The Sobel standard error is implemented in several statistical software programs, such as LISREL, EQS, and *Mplus*, and will be used in all three studies comprising this dissertation.

It is important to note that the Sobel standard error is not without its critics. The use of the Sobel standard error relies on asymptotic theory (i.e., normal-theory confidence limits and hypothesis tests). The distribution of a product of two independent normally distributed random variables is generally not normally distributed, although it may approach normality in large samples (MacKinnon, 2008a). The net result is generally conservative hypothesis tests (low Type I error rates and low power) and confidence intervals (i.e., empirical coverage probabilities larger than 95% for a 95% confidence interval). An alternative to the Sobel test is the joint significance test of mediation discussed by Cohen and Cohen (1983). This test declares a significant mediated effect when the paths comprising the mediated effect are *both* statistically significant. For the basic single-mediator model in Figure 1, two null hypotheses are

tested ($H_0 : \alpha = 0$ and $H_0 : \beta = 0$) and there is evidence for mediation when both null hypotheses are rejected (i.e., the paths are *jointly* significant). No estimate of the mediated effect is necessary and such a test does not provide confidence intervals.

Other alternatives to the Sobel standard error (and similarly calculated measures) include one of several methods of bootstrapping (Shrout & Bolger, 2002; MacKinnon, Lockwood, & Williams, 2004; Preacher & Hayes, 2004; Cheung & Lau, 2008) or methods that construct asymmetric confidence intervals based on the theoretical distribution of the product of two random normal variables (hypotheses tests can be conducted based on such confidence intervals) (MacKinnon et al., 2004, MacKinnon et al., 2007).

Some have advocated for the use of a proportion mediated measure as an alternative to the product of coefficients and difference in coefficients methods to quantify mediation (e.g., see Vittinghoff, Sen, & McCulloch, 2009). In the case of a continuous outcome variable with no missing data, the following formulae provide equivalent values of the proportion mediated effect:

$$\frac{\tau - \tau'}{\tau}, \quad \frac{\alpha\beta}{\tau}, \quad \frac{\alpha\beta}{\alpha\beta + \tau'}$$

It is important to note that proportion mediated effect can be less than zero or greater than one. In addition, based on simulation studies, the stability of this measure has been called into question unless the effect is large or the sample size is large ($N > 500$) (MacKinnon, Warsi, & Dwyer, 1995).

1.1.3 Mediation Modeling with Cross-Sectional Data

As defined earlier, mediation implies a temporal relation among X, M, and Y (MacKinnon, 2008a). Thus, X should occur before M and M should occur before Y. This temporal relationship may help to clarify differences between mediation and other third-variable effects. For example, time-invariant variables (e.g., gender) might be conceptualized and shown to serve as moderators (i.e., effect modifiers), but they cannot be intermediate variables because an independent variable cannot change a time-invariant variable. More importantly though, it helps shed some light on the difficulties associated with establishing mediation relationships with cross-sectional data.

MacKinnon (2008b) describes several benefits of using longitudinal data to evaluate mediation processes. First, more information regarding the temporal sequentiality of X, M, and Y is provided with longitudinal data, a critical underlying assumption of mediation. Second, longitudinal data allow for an examination of associations within waves of data (i.e., cross-sectional, between subjects) and changes across waves of data (i.e., within individuals). Finally, individuals may serve as their own control in assessing some relationships, thereby potentially controlling for static differences among individuals.

In their discussion of why cross-sectional data generally provide poor estimates of effects, Gollob and Reichardt (1991) note another related set of benefits to the use of longitudinal data to assess mediation. First, causal effects often take time to develop, and variables measured at the same time may not allow for the necessary development time. Second, variables often have effects on themselves, such that Y at a later time is related to Y at an earlier time (cross-sectional models implicitly assume these autoregressive

effects are zero). Related to this issue, Cole and Maxwell (2003) note that it is not sufficient to merely allow a time lag between X and M and then between M and Y to achieve unbiased estimates of effects because of the potentially confounding effects of prior levels of the M and Y (for example of a published article using such an approach to longitudinal mediation analysis, see Plotnikoff, Pickering, Flaman, & Spence, 2010). Third, effect sizes often depend on the amount of time separating observations. While this may often present itself as a considerable problem with longitudinal studies as will be discussed later (Gollob & Reichardt, 1991; Collins & Graham, 2002; Cole & Maxwell, 2003), cross-sectional data ignore such potentially problematic issues by simply not specifying the time interval under study.

These issues suggest that the use of cross-sectional data will often result in biased estimates of true causal effects. In the case of true longitudinal mediation, Maxwell and Cole (2007) demonstrate that cross-sectional analyses provide biased estimates of the indirect effect in two different models of change, the autoregressive model and a random effects model (their work assumes continuous X, Y, and M that all change over time, as well as the case of complete mediation). These models will be reviewed in a subsequent section of this paper (and variants of each will be evaluated in parts of this dissertation). The pattern of bias (i.e., direction and degree) for these two models depends on different factors. For example, the pattern is strongly related to the degree of stability (defined in their work as the correlation of a variable with itself at time t and time $t + 1$) in X and M in the autoregressive model and to the extent of variability across individuals in baseline levels of M and Y, as well as the degree of correlation between these baseline levels with

each other and with levels of X (essentially covariance among random effects), in the random effects model.

Gollob and Reichardt (1991) propose and develop a latent longitudinal model for use with cross-sectional data (the “missing” measures at earlier time points are not measured, hence the term latent). Several constraints must be added in order for their model to be identified and some have argued that these constraints (assumptions) are often unrealistic, and testable only with longitudinal data (Cole & Maxwell, 2003; MacKinnon, 2008b).

This brief review suggests that in many situations there may not be an adequate substitute for longitudinal data when the goal is to assess mediation. Given this, several different options for modeling mediation with multiple waves of data have been developed. These models are based on techniques developed for longitudinal data analysis in general (i.e., without the explicit goal of assessing mediation). Although these models have some well understood properties, there is still a significant amount of unknown information concerning the appropriate use of the models and the conditions under which these models accurately represent true population effects.

1.1.4 Summary of Basic Mediation Modeling

Before exploring models for assessing longitudinal mediation, it may be helpful to summarize issues with respect to basic mediation modeling. Numerous articles estimating and testing mediated effects, especially in the psychology literature, have been published in a variety of substantive areas, most typically employing a cross-sectional design with X, M, and Y measured simultaneously, usually as continuous variables. In

addition to these content applications, a number of methodological and statistical issues with respect to mediation have been explored. Many of these issues have been investigated with Monte Carlo simulation methods (Paxton, Curran, Bollen, Kirby, & Chen, 2001; Fan & Fan, 2005; Bandalos, 2006) while others have been investigated analytically (e.g., Maxwell & Cole, 2007). For example, authors have investigated the effects of nonnormality in the estimating of mediated effects (Finch, West, & MacKinnon, 1997), while several other authors have explored the impact of measurement error on estimates and tests of mediation in the cross-sectional mediation model (Hoyle & Kenny, 1999; Stephenson & Holbert, 2003; Cheung & Lau, 2008). Several studies have examined the accuracy of various standard error formulae for the mediated effect (e.g., MacKinnon et al., 2002), while others have examined the related topic of confidence interval construction, including the use of resampling methods such as bootstrapping (Shrout & Bolger, 2002; Preacher & Hayes, 2004; MacKinnon, Lockwood, & Williams, 2004; MacKinnon et al., 2007; Cheung 2007; Cheung & Lau, 2008). Still others have explicitly considered confidence intervals for the *standardized* indirect effect (Cheung, 2009). Raykov and colleagues (2008) developed a general strategy for comparing the sizes of mediated effects.

Multiple mediator models, both the situation when multiple variables mediate the relationship between X and Y (i.e., $X \rightarrow M_1 \rightarrow Y$ and $X \rightarrow M_2 \rightarrow Y$) (MacKinnon, 2008c) as well as the situation of longer chain mediation models ($X \rightarrow M_1 \rightarrow M_2 \rightarrow Y$) (Taylor, MacKinnon, & Tein, 2008) have been explored. The concepts of mediated moderation and moderated mediation were described by Baron & Kenny (1986). Several authors have recently discussed procedures to address these concepts (Muller, Judd, & Yzerbyt,

2005; Edwards & Lambert, 2007; Preacher, Rucker, & Hayes, 2007; Fairchild & MacKinnon, 2009).

Finally, ubiquitous in its assessment in Monte Carlo simulation studies of mediation models (and structural equation modeling more generally) is the role of sample size. This is because many applications routinely use small sample sizes and the impact on the properties of point estimators, confidence intervals, hypothesis tests, and measures of goodness of fit is often not well understood for small to moderate sample sizes (Paxton et al., 2001).

Mediation continues to be a very active area of methodological and statistical research, primarily in psychology, but in other areas as well. For example, research concerning sample size calculations for mediation analysis has been published in both psychology (Fritz & MacKinnon, 2007) and statistics (Vittinghoff, Sen, & McCulloch, 2009) journals.

Despite the large number of application-based and methods-based research conducted in the area of mediation modeling, numerous issues remain unresolved. For example, causal inference approaches for evaluating mediation models are receiving significant attention in the literature (see MacKinnon, 2008d, for a review). Expansion of the general mediation model to different types of variables, such as count variables and two-part models (used when there are a preponderance of zeros, such as the modeling of health expenditures (see Dow and Norton, 2003)), remain largely unexplored. Finally, despite the availability of several models for exploring longitudinal mediation, a number of issues regarding the evaluation of these models remain unexplored. For example, several years ago Cheong, MacKinnon, and Khoo (2003, p. 260) observed the following

regarding one specific model for exploring longitudinal mediation which still holds today:

Although applications of parallel process LGM [latent growth modeling] have begun to appear in the substantive research, little is known about how accurately the parameters are recovered. Researchers can set the parameter values of the population model for the parallel process and investigate the conditions under which an accurate representation of the true population is reproduced. In addition, Type I error rates and statistical power can be investigated.

1.2 Longitudinal Mediation Modeling

Before discussing the specific longitudinal mediation models that will be explored in this dissertation (i.e., models for which three or more waves of data are collected), it may be useful to review some terminology and to also briefly mention available options for two-wave models. The concepts of stability, stationarity, and equilibrium are frequently mentioned in discussions of longitudinal models (Kenny, 1979; Cole & Maxwell, 2003; Maxwell & Cole, 2007; MacKinnon, 2008b). Although *stability* is sometimes referred to as an unchanging mean of a variable over time (MacKinnon, 2008b), it is also used to refer to a substantial correlation between the same variable measured at time t and $t + 1$. *Stationarity* refers to whether the relationships among variables are the same over time (i.e., an “unchanging causal structure”) while *equilibrium* suggests the cross-sectional variances and covariances among variables do not depend on the time of measurement. In order to assume that the observed relationships among X, M, and Y are not simply due to the time of their measurement, some degree of equilibrium and stationarity must be present. Fortunately, these assumptions can be tested with longitudinal data. However, they cannot be tested with

cross-sectional data and thus must be (sometimes fallaciously) assumed when cross-sectional data are used to test mediation (Cole & Maxwell, 2003). In addition, the degree of stability in X and M plays a large role in determining the pattern of bias when cross-sectional data are used to estimate longitudinal mediation.

Although some advocate that three waves of data are necessary to truly test mediation (Collins, Graham, & Flaherty, 1998), it is possible to provide some evidence of mediation with two waves of data. MacKinnon (2008b) outlines three approaches:

1. Difference scores – Difference scores ($t_2 - t_1$, where t_i = measurement at time $i = 1, 2$) are calculated for each of X, M, and Y and these difference scores are then used in the traditional mediation model (Figure 1).
2. Residualized change scores – Predicted scores for t_2 are calculated based on t_1 scores (using the linear regression of t_2 on t_1) for each of X, M, and Y. The residualized change score is the difference between t_2 observed scores and predicted scores at t_2 (based on t_1). The residualized change scores for X, M, and Y are then used in the traditional mediation model (Figure 1).
3. ANCOVA – A variety of different models are available in which t_1 measures are used as covariates in models with either raw scores at t_2 or change scores used as dependent variables. One possible path diagram can be found in Figure 2 (Cole & Maxwell, 2003). This model is called a two-wave autoregressive model by MacKinnon (2008b). Assuming stationarity, $a*b$ provides an estimate of the mediated effect of X on Y through M. MacKinnon (2008b) elaborates on this model and provides estimates of contemporaneous effects as well (i.e., $X_2 \rightarrow M_2$, and $M_2 \rightarrow Y_2$). It should be noted that $a*b$ is not usually equal to $c - c'$ (this is the

difference in coefficients estimator of the mediated effect described in section 1.1.2, where c and c' are used instead of τ and τ' , respectively) in two-wave models and covariate-adjusted wave 2 effects are not generally equivalent to covariate-adjusted change score effects, introducing some uncertainty and variability in methods for testing two-wave mediation effects (Roth & MacKinnon, 2012).

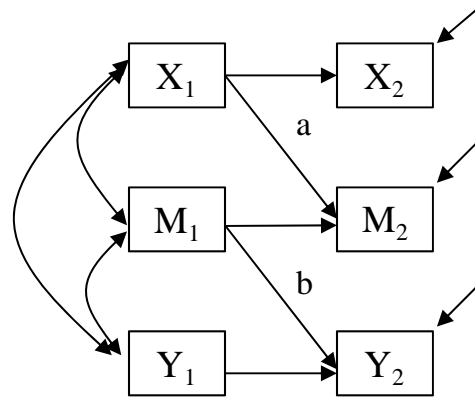


Figure 2. Two-wave autoregressive model.

The increased complexity in two-wave models over cross-sectional models is extended and becomes even more complex in models with three or more waves of data. Yet, such models offer much promise for providing for precise and compelling tests of mediation hypotheses. Longitudinal models with three or more waves of data comprise the over-riding focus of this dissertation.

There are several classes of models for evaluating longitudinal mediation with the collection of three or more waves of data, including 1) autoregressive models, sometimes called panel models or cross-lagged panel models (Gollob & Reichardt, 1991; Cole & Maxwell, 2003), 2) different types of random effects models (Kenny, Korchmaros, & Bolger, 2003; Cheong, MacKinnon, & Khoo, 2003; Cheong, 2011; von Soest & Hagtvet,

2011), 3) a stage-sequential model of mediation based on binary data (Collins, Graham, & Flaherty, 1998), 4) adaptations of latent difference score models which are useful when electing to examine potential differences at different waves of data (McArdle, 2009; Geiser et al., 2010), and 5) an autoregressive latent trajectory (ALT) model, which combines elements of 1 and 2 (Curran & Bollen, 2001). Because versions of autoregressive and random effects models will be explored in this dissertation, we will further explore these models.

1.2.1 Autoregressive Mediation Models

The basic premise of autoregressive models is that the values of a variable at a future time point depend in some part on an earlier time point. The simplest of these models involves one variable measured at multiple time points for a set of individuals, where the value of the variable (Y) for individual i at time $t + 1$ is a linear function of that individual's value of Y at time t :

$$Y_{it+1} = \beta_0 + \beta_1 Y_{it} + \varepsilon_{it+1} \quad (1.1)$$

where β_1 is an index of stability over time.

This basic model (the univariate model) is often referred to the simplex model (or the Markov simplex model) (Marsh, 1993; Curran & Bollen, 2001) and dates back to the work of Guttman (1954). Gollob and Reichardt (1991) and subsequently Cole and Maxwell (2003) extended the univariate simplex model to a trivariate model, an autoregressive mediation model, involving X, M, and Y (see Figure 3 for the basic structure of this model with three waves – note that in this model a is used instead of α , b instead of β , and c' instead of τ' to correspond to the notation used by Cole & Maxwell

(2003), Maxwell & Cole (2007) and MacKinnon (2008b)). Parameter estimation and the calculation of standard errors involve solving a system of simultaneous equations and is usually accomplished using covariance structure analysis (also called structural equation modeling – SEM) (Cole & Maxwell, 2003; MacKinnon, 2008b). In essence, this involves finding a set parameter estimates that yields a predicted variance-covariance matrix (one that is implied by the estimates of the unknowns) that most closely reproduces the sample variance-covariance matrix. The traditional method of estimation in SEM is maximum likelihood, although many other options are available and are often the focus of Monte Carlo simulation studies in SEM (Gerbing & Anderson, 1993; Paxton et al., 2001; Fan & Fan, 2005; Bandalos, 2006).

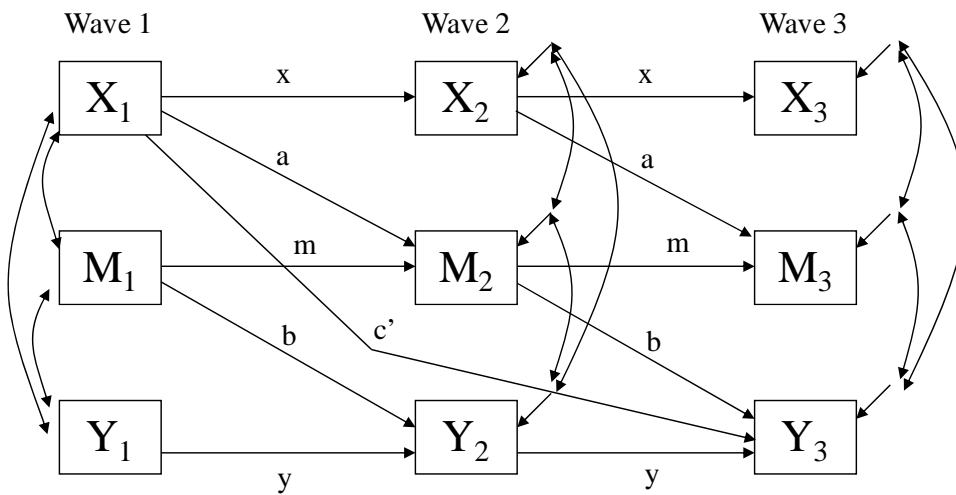


Figure 3. Basic three-wave autoregressive mediation model.

The basic longitudinal mediated effect in the model in Figure 3 is $a*b$. Assuming stationarity, which is testable in this model, it should not matter which a or b is used, however it is common to use $X_1 \rightarrow M_2$ as the a path and $M_2 \rightarrow Y_3$ as the b path to reflect temporal ordering. In a single-mediator, cross-sectional model of mediation, the condition of a zero direct effect (i.e., the relationship between X and Y adjusted for M)

implies the case of complete or perfect mediation (Baron & Kenny, 1986). A direct effect of X_1 on Y_3 is represented by the c' path in Figure 3. Although the condition of $c' = 0$ is necessary, it is not a sufficient condition for mediation to be complete in the autoregressive mediation model. Assuming the x and y paths are nonzero, X_1 could directly affect Y_3 (i.e., not go through M) through Y_2 (i.e., $X_1 \rightarrow Y_2 \rightarrow Y_3$ is a time-specific direct effect). In autoregressive mediation models, these other direct effects must also be zero for mediation to be complete (Cole & Maxwell, 2003).

The earlier discussion with respect to the calculation of standard errors and the construction of confidence intervals in the traditional (i.e., cross-sectional) mediation model still applies (MacKinnon, 2008b). It is worth briefly exploring the covariance among the residuals at waves 2 and 3. MacKinnon (2008b) suggests that this is necessary to reflect that there are contemporaneous relationships among X , M , and Y , but the causal order is unknown. A more informative explanation is provided by Cole and Maxwell (2003, p. 571), who note that the finding of significant and meaningful covariation among these residuals implies that “potentially important variables are missing from the model.” These variables may be important confounders that, if ignored, may lead to biased estimates.

There are numerous extensions to the model in Figure 3. Lag 1 direct effects of X to Y can be added ($X_1 \rightarrow Y_2$; $X_2 \rightarrow Y_3$), implying additional mechanisms of partial mediation. Contemporaneous mediation relationships might be tested (i.e., $X_2 \rightarrow M_2 \rightarrow Y_2$). The presence of “theoretically backward effects” (Cole & Maxwell, 2003, p. 571) or cross-lagged relationships can be tested (i.e., $Y_1 \rightarrow M_2$, $M_2 \rightarrow X_3$). The addition of more waves allows one to estimate both time-specific indirect effects (i.e.,

$X_1 \rightarrow M_2 \rightarrow Y_3$), as well as the overall indirect effect (Gollob & Reichardt, 1991; Cole & Maxwell, 2003). For example, for the four wave model in Figure 4, there are three time-specific indirect effects:

1. $X_1 \rightarrow X_2 \rightarrow M_3 \rightarrow Y_4$
2. $X_1 \rightarrow M_2 \rightarrow M_3 \rightarrow Y_4$
3. $X_1 \rightarrow M_2 \rightarrow Y_3 \rightarrow Y_4$

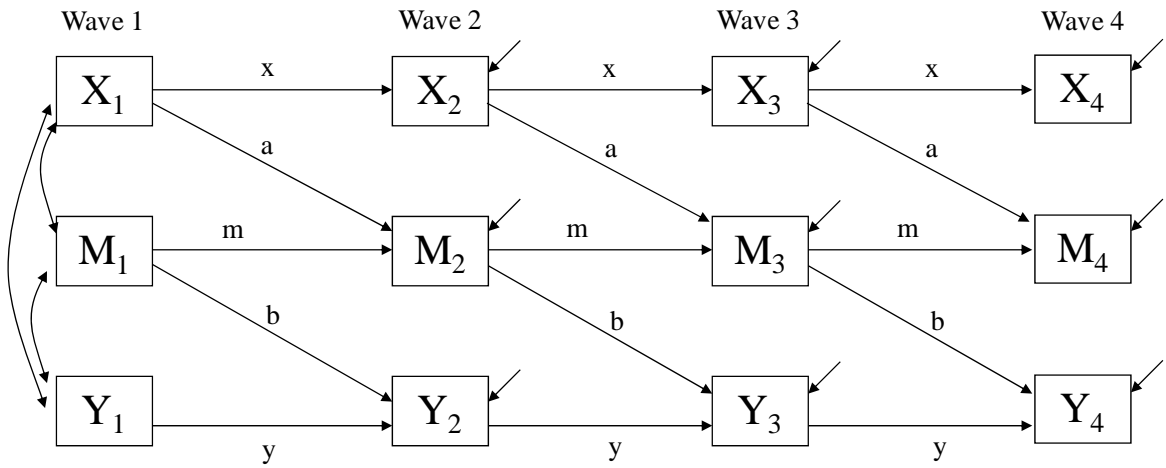


Figure 4. Four-wave autoregressive mediation model. Note that several arrows have been removed to simplify the figure.

The overall indirect effect of $X_1 \rightarrow Y_4$ is the sum of the time-specific indirect effects. In this case: $abx + abm + aby$. This addresses the question of whether M mediates the effect of X_1 on Y_4 at any time between waves 1 and 4 (rather than at some specific point), a question usually of greater interest to researchers (Cole & Maxwell, 2003). More waves of data create a greater number of indirect effects from which to choose. It also creates issues with respect to standard error calculations. Taylor, MacKinnon, and Tein (2008) have derived formulas for standard errors for the three-path mediated model, extending the work of Sobel (1982) and Goodman (1960), but these calculations are complex and grow in complexity with more waves (and may be quite

inappropriate for smaller samples due to the reliance on asymptotic theory).

Furthermore, the distributional theory of complex indirect effects (i.e., the overall indirect effect) generated from multiple wave panel designs is still largely unexamined (Little et al., 2007), generally necessitating the use of bootstrapping.

Although the autoregressive mediation model certainly has advantages over the use of cross-sectional mediation modeling, there are several limitations and challenges associated with its use. The number of possible indirect effects, especially with more than three waves of data and with the addition of contemporaneous mediated effects, has already been discussed. As with any longitudinal modeling effort, missing data can create problems, especially if the missing data pattern is non-ignorable. In addition, the timing and spacing of measurements can dramatically influence the detection and estimation of effects in longitudinal studies (Collins and Graham, 2002). Furthermore, it is possible that the timing of assessments to observe the maximal treatment effect may not be the same as the timing to observe mediated effects (Gollob & Reichardt, 1991; Cole & Maxwell, 2003). Autoregressive models do not include modeling of means, nor do they allow for random effects to specify individual differences in change (MacKinnon, Fairchild, & Fritz, 2007).

As with the case of cross-sectional models of mediation, the impact of measurement error in X, M, and Y, can be substantial and potentially more complex (Cole & Maxwell, 2003), as will be discussed in paper two. MacKinnon (2008, p. 209) notes, “One way to improve the interpretability of autoregressive models is to improve measurement of variables either by specifying latent variables or increasing the reliability of measures.” The autoregressive mediation model can be extended to latent variable

models, where latent variables are extracted from multiple indicators (i.e., observed variables) for X, M, or Y (see Roth, Mittelman, Clay, Madan, & Haley (2005) for an example of a two-wave model that uses latent variables for the mediators). As will be discussed later, this is a commonly used method to address the biasing effect of measurement error, but introduces other modeling options that should be evaluated and compared in methodological research. In addition, models with latent variables allow researchers to test whether there is measurement invariance over time (Brown, 2006). In other words, researchers can assess whether changes over time in a measure can be assumed to be true score changes or whether the meaning of the measure itself has changed over time. While the use of latent variables certainly has its benefits, the use of latent variables increases the possibilities of model misspecification, including the failure to account for shared method variance, which may lead to biased parameter estimates (Marsh, 1993; Cole, Ciesla, & Steiger; 2007; Geiser et al., 2010; Kline, 2011).

1.2.2 Random Effects Mediation Models

There are two ways to conceptualize random effects mediation models, the primary difference being whether or not time is explicitly incorporated into the model as a variable. Before exploring these models, it is worth clarifying some terminology. Researchers from the educational, social, and behavioral sciences often use the terminology of hierarchical linear models (HLM) or multilevel models (MLM), whereas statisticians often use terminology of the linear mixed model. The models are actually the same, although simply different in form (i.e., the combined or composite model from HLM is in the linear mixed model form) (Feng et al., 2001; Singer & Willett, 2003).

Although it will not be used in this dissertation, it may be helpful to briefly review the random effects model of mediation proposed by Kenny, Korchmaros, and Bolger (2003), which does not explicitly incorporate time as a variable. In their example, which was also used as the basis for the work of Maxwell and Cole (2007), all three variables are level-1 variables, or time-varying variables when applied to the repeated measures (i.e., the upper level or level-2 units are individuals and the lower level or level-1 units are the repeated measures). Hence, the authors refer to their model as a lower-level mediation model (Krull & MacKinnon (2001) label the model $1 \rightarrow 1 \rightarrow 1$, to signify all variables are at the lower level). In their example, X is daily stress level, M is coping efforts for that day, and Y is the person's mood for that day. The following equations (notation based on Maxwell and Cole (2007) and Kenny, Korchmaros, and Bolger (2003)) are used to evaluate mediation (notice that time is not a variable in the model, but is implicitly included as X, M, and Y are allowed to vary over time – i.e., are time varying variables):

$$\begin{aligned}
 &\text{Level 1 equation: } M_{it} = d_{1i} + a_i X_{it} + e_{it} \\
 &\text{Level 2 equations:} \\
 &d_{1i} = d_1 + u_{1i} \\
 &a_i = a + u_{2i}
 \end{aligned} \tag{1.2}$$

$$\begin{aligned}
 &\text{Level 1 equation: } Y_{it} = d_{2i} + c'_i X_{it} + b_i M_{it} + f_{it} \\
 &\text{Level 2 equations:} \\
 &d_{2i} = d_2 + u_{3i} \\
 &b_i = b + u_{4i} \\
 &c'_i = c' + u_{5i}
 \end{aligned} \tag{1.3}$$

where X_{it} , M_{it} , and Y_{it} represent values of X, M, and Y for individual i at time t and M_{it} and is a linear function of an individual's X value on a given occasion (X_{it}) represented

by a_i and Y_{it} is a linear function of both an individual's X and M values (X_{it} and M_{it}) on a given occasion, represented by c'_i and b_i , respectively.

An examination of equations 1.2 and 1.3 shows that the effects of interest, a_i , b_i , and c'_i , can vary across individuals (i.e., they are random). The mean of the random effects a_i and b_i (i.e., a and b in equations 1.2 and 1.3, respectively; also known as the fixed effects), can be used to estimate the mediated effects through the product of coefficients method. However, because the relation of X to M (coefficient a_i) and the relation of M to Y (coefficient b_i) can vary across individuals (i.e., are random), the covariance between a_i and b_i must be added to $a*b$ to arrive at an unbiased estimate of the mediated effect (Kenny, Korchmaros, & Bolger, 2003; MacKinnon, 2008e). A similar consideration must be made in the calculation of the standard error. The necessary estimates can be arrived at by fitting separate equations in a mixed model program (e.g., PROC MIXED in SAS) and then use the procedures outlined by Kenny, Korchmaros, and Bolger (2003) to estimate the covariance between a_i and b_i . One may also fit the entire set of equations using a program such as *Mplus* (see MacKinnon, 2008e).

The other random effects model of mediation explicitly incorporates time as a variable. The basic model on which the mediation framework is based is a type of longitudinal data modeling (i.e., without the explicit goal of assessing mediation) that goes by a variety of names, including the multilevel model for change (Singer & Willett, 2003) and latent growth curve (LGC) modeling (Duncan & Duncan, 2004). The basic multilevel model for change (i.e., without mediators or any other predictors – the unconditional growth model of Singer and Willett (2003)) is:

Level 1 equation: $Y_{it} = \pi_{0i} + \pi_{1i}TIME_{it} + \varepsilon_{it}$

Level 2 equations:

$$\pi_{0i} = \gamma_{00} + \zeta_{0i}$$

$$\pi_{1i} = \gamma_{10} + \zeta_{1i}$$

(1.4)

where Y_{it} represents the value of Y for individual i at time t and $TIME_{it}$ indicates the measurement occasion (i.e., some sensible metric that indicates the passage of time) – often assumed to be a linear growth trajectory, but other patterns are certainly possible.

Thus, in equation 1.4, Y_{it} is a linear function of time and the intercept π_{0i} and the effect of time (i.e., the slope, π_{1i}) can vary across individuals (i.e., they are random). The mean of these random variables, γ_{00} and γ_{10} , are the fixed effects in this model. Thus, different individuals can have their own growth pattern (both initial starting value and slope).

Time-varying predictors can be added at level 1 or predictors of the coefficients for the intercept (initial status) or time can be added at level 2 (i.e., time-invariant predictors) to indicate that the growth varies as a function of the time-invariant predictor (e.g., assignment to treatment condition). The parameters from such models can be estimated in a mixed model program (e.g., PROC MIXED in SAS).

The basic multilevel model for change also can be constructed as a structural equation model (SEM). This analytical approach has come to be known as latent growth curve modeling. In a SEM framework, the growth of a variable over time can be represented in a measurement model in matrix form as:

$$\text{Growth measurement model for a single individual: } \mathbf{Y}_i = \mathbf{\Lambda}\boldsymbol{\eta}_i + \boldsymbol{\varepsilon}_i \quad (1.5)$$

As noted by Cheong, MacKinnon, and Khoo (2003), \mathbf{Y}_i is a $T \times 1$ vector of repeated measures of the variable Y for individual i over the time points ($t = 0, 1, 2, \dots, T$), $\mathbf{\Lambda}$ is a $T \times J$ matrix of factor loadings on the growth factors, $\boldsymbol{\eta}_i$ is a $J \times 1$ vector of J latent factors

representing the growth parameters, and $\mathbf{\epsilon}_i$ is a $T \times 1$ vector of measurement errors. A conceptual representation of this model where $T = 3$ (three measurement occasions) and $J = 2$ (two growth parameters) can be found in Figure 5.

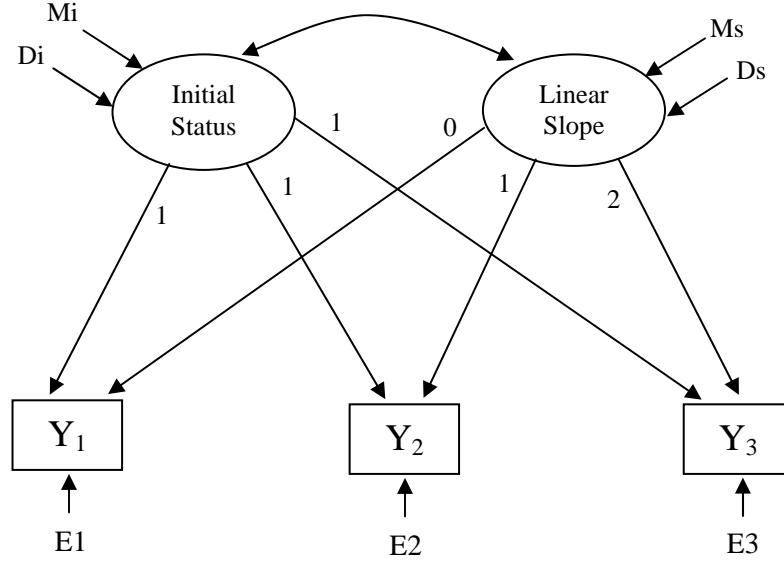


Figure 5. Basic three-wave latent growth curve model (M_i and D_i are the mean and variance of the initial status factor while M_s and D_s are the mean and variance of the linear slope factor).

It is possible to specify exactly the same model as a LGC model or a multilevel model for change (Hox & Stoel, 2005).² The TIME variable is incorporated into the LGC model as specific constrained values for the factor loadings in the Λ matrix: values for the loadings for the initial status factor η_{1i} are constrained to be 1 and values for the loadings of the growth factor η_{2i} are fixed by the analyst to reflect time intervals between measurements and the shape of the growth trajectory (e.g., linear, quadratic, etc.). For example, in the three-wave LGC model in Figure 5, the loadings of [0 1 2] for η_{2i} reflect a linear trajectory across 3 time points at evenly spaced intervals (hence that factor is

²To make the models exactly equivalent, one must constrain the variances of E1-E3 in Figure 5 to be equal.

labeled “linear slope”). One can estimate the mean of the initial status factor (or the intercept factor), which corresponds to the fixed effect, γ_{00} , in the multilevel model for change. In addition, the mean of the growth factor corresponds to the fixed effect, γ_{10} . Individual differences in parameters describing the growth curve in the LGC model are modeled as (co)variances of the intercept (initial status) and growth (or slope) factors. In essence, these correspond to the variances and covariance of ζ_{0i} and ζ_{1i} in the multilevel model for change (the random effects). The primary advantage of using a LGC modeling approach rather than mixed modeling methods (i.e., the multilevel model for change) is the ability to examine several response variables within the context of a single latent variable model (Duncan & Duncan, 2004). This distinction is no longer that meaningful as several researchers have shown it is possible to include multiple response variables using an extended version of the multilevel model for change (i.e., multivariate longitudinal data) (Thiébaud, Jacqmin-Gadda, Chêne, Leport, & Commenges, 2002; Gao, Thompson, Xiong, & Miller, 2006).

The multilevel model for change and the latent growth curve model can be expanded to include mediation processes (i.e., this is the other random effects model of mediation that explicitly incorporates time as a variable). Perhaps the most straightforward approach is to model growth factors based on repeated measures of X, M, and Y. With this approach, called a parallel process model (MacKinnon, 2008b), one can assess mediation by exploring the relationships in growth among X, M, and Y. One possible parallel process model can be found in Figure 6. Note that the mediated effect in Figure 6, $a*b$, is comprised of two random effects (these relationships can vary across

participants), thus the point estimate and standard error calculations must take this into account (see Kenny, Korchmaros, & Bolger, 2003; MacKinnon, 2008b).

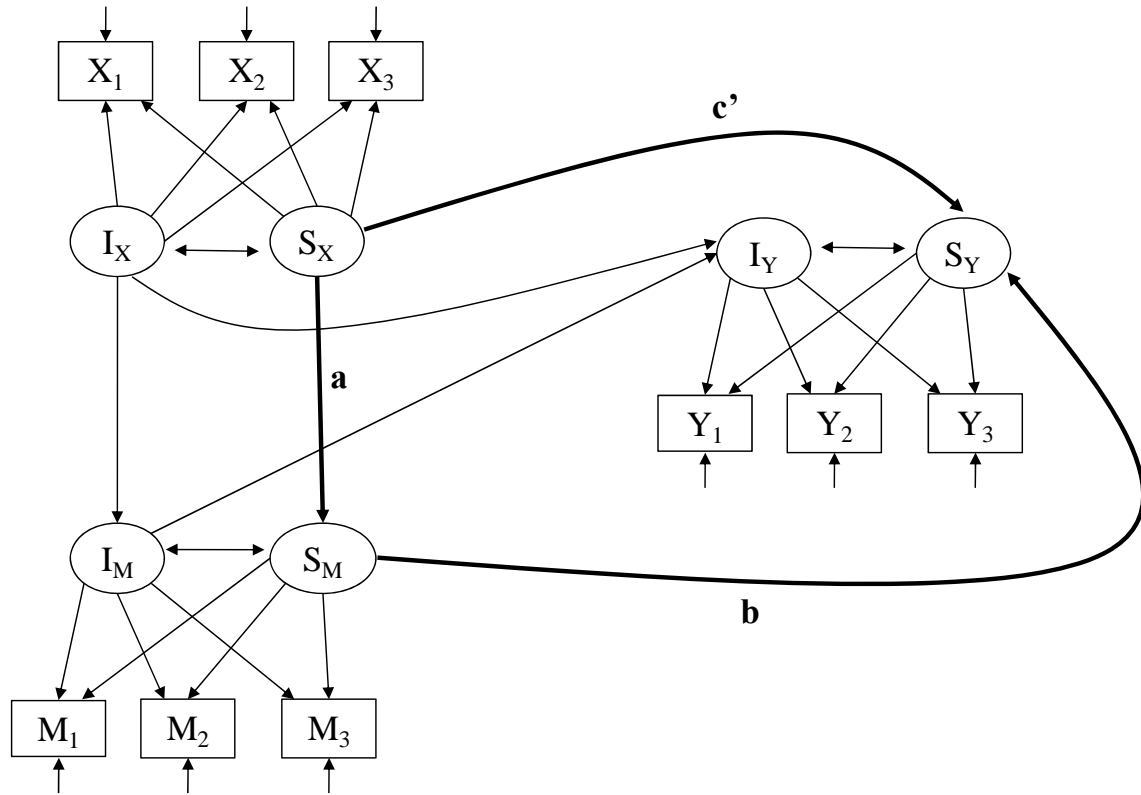


Figure 6. Three parallel process latent growth curve mediation model with three waves of data for the independent variable (X), the mediator (M), and the dependent variable (Y) (adapted from MacKinnon, 2008b). I_X represents the intercept factor (i.e., initial status) of the independent variable, S_X represents the slope (growth rate) factor of the independent variable, I_M represents the intercept factor (i.e., initial status) of the mediator, S_M represents the slope (growth rate) factor of the mediator, I_Y represents the intercept factor (i.e., initial status) of the outcome, and S_Y represents the slope (growth rate) factor of the outcome. The mediated effect is represented as $a*b$ and the direct effect is c' . Note that fixed factor loadings specifying the initial status and slope factors are not shown to simplify the figure. Also note additional paths may be added as needed, such as $I_X \rightarrow S_M$, $I_X \rightarrow S_Y$, $I_M \rightarrow S_Y$, and $I_Y \rightarrow S_M$. One may choose to specify covariances rather than directed paths among some intercepts and slopes to show that there are relationships among the variables, but directions are unknown.

A related LGC mediation model was developed by Cheong, MacKinnon, and Khoo (2003) for the situation where X is a fixed intervention rather than a time-varying variable, like M and Y (Figure 7). In this model, the $X \rightarrow M$ path (the “a” path) is a fixed effect, therefore standard approaches to estimating the point estimate and standard error for the mediated effect can be used (Cheong, MacKinnon, & Khoo, 2003). Cheong, MacKinnon, and Khoo (2003) outline several steps for evaluating parallel process mediation models. This includes evaluating both M and Y for differential growth trajectories, using multiple-groups SEM to evaluate differences in initial status and growth across the levels of the X (e.g., see Kline, 2011), combining the latent growth curve models for M and Y into one parallel process model and then using MIMIC (multiple indicators, multiple causes) modeling (e.g., see Brown, 2006) to assess the impact of the treatment (X) on the growth factors, and finally adding relationships among the growth factors to assess for mediation.

As with the case of the autoregressive mediation model, there are some limitations and challenges associated with the use of the LGC mediation model. Since the LGC methodology is a type of longitudinal modeling, non-ignorable missing data patterns can create problems. Although the timing and spacing of measurements may be less of an issue for LGC models than autoregressive models if the growth trajectory is represented well by a straight line (Collins & Graham, 2002), it can still cause some problems, especially if not enough time has elapsed in order for effects to materialize. Measurement, including measurement error and invariance in the meaning of a measure over time, can have substantial impacts on the interpretation of the results, just like in the autoregressive mediation model. A common criticism of the LGC mediation model is

that the mediation relationship is correlational in nature (i.e., there is no temporal precedence of the mediator as there is in the autoregressive mediation model). Although the presence of a temporal sequence alone does not unequivocally establish causality, such a sequence is considered stronger evidence for a causal relationship than a simultaneous relationship (Hill, 1965; Kazdin & Nock, 2003). Several authors note the possibility of using a two-stage piecewise parallel process approach to overcome this criticism. In this model, the growth in the mediator and outcome can be modeled as occurring in separate phases. Thus, the mediated effect can be evaluated via earlier growth of the mediator on later growth in the outcome (Figure 8) (Cheong, MacKinnon, & Khoo, 2003; Laurenceau, Hayes, & Feldman, 2007; McKinnon, 2008; Cerin, 2010). An elaboration of this model is provided in paper three. Although the putative mediator in the parallel process models depicted in Figures 7 and 8 is the slope (or growth rate), it is possible to test mediated effects involving the mediator's intercept factor (von Soest & Hagtvet, 2011).

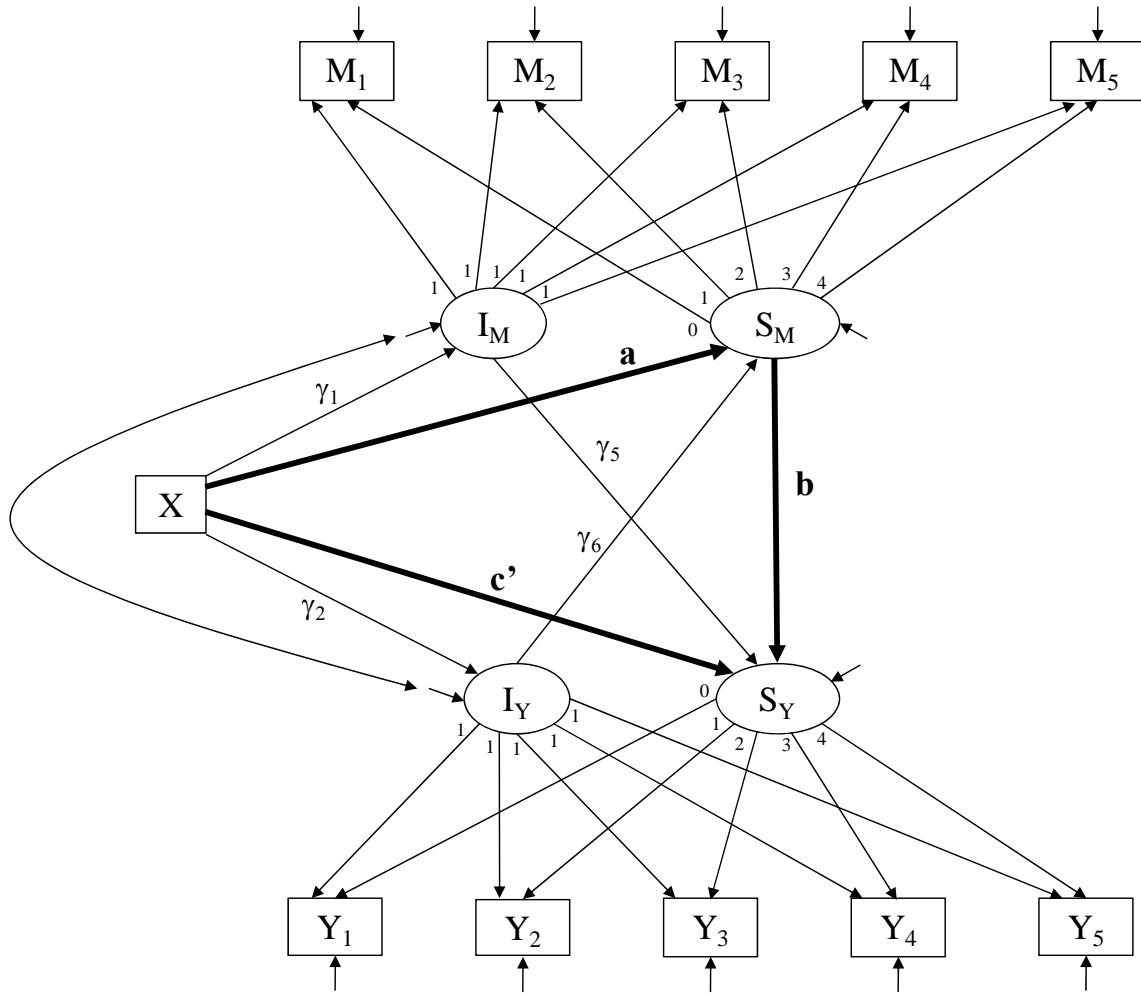
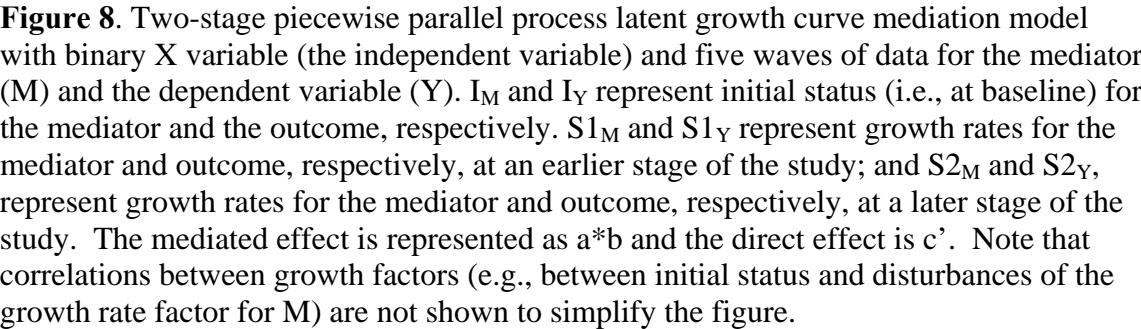


Figure 7. Parallel process latent growth curve mediation model with binary X variable (the independent variable) and five waves of data for the mediator (M) and the dependent variable (Y) (adapted from Cheong et al., 2003). I_M represents the intercept factor (i.e., initial status) of the mediator, S_M represents the slope (growth rate) factor of the mediator, I_Y represents the intercept factor (i.e., initial status) of the outcome, and S_Y represents the slope (growth rate) factor of the outcome. The mediated effect is represented as $a*b$ and the direct effect is c' . Note that correlations between growth factors (e.g., between initial status and disturbances of slope factor for M) are not shown to simplify the figure.



1.2.3 Comparison of Longitudinal Mediation Models

A natural question regarding the assessment of longitudinal mediation is which model to use. Unfortunately, there is not a simple answer to this question. There have been some studies and manuscripts in which autoregressive and latent growth curve models have been compared, usually in the context of a single variable measured over time (i.e., studies without mediation). Perhaps the most widely cited is Rogosa and Willett (1985), who showed that a quasi-simplex model (a simplex model that allows for observed variables with measurement error) fits well to data generated from a growth model (in this case, a constant rate of change model, which is essentially a latent growth curve model with initial status and linear slope factors). The authors used these results to question the use of the simplex model to study growth as such models may not be able to distinguish among different patterns of individual growth. Subsequent authors have provided examples to the contrary, showing evidence that quasi-simplex models do not fit data generated from other types of growth models (Mandys, Dolan, & Molenaar, 1994; Raykov, 2000). Others have shown that Rogosa and Willett's (1985) findings are partially based on the covariance matrix they chose to analyze (Raykov, 1998).

Several have noted that autoregressive models, including simplex models, are group-change or group-coefficient models, whereas LGC models are individual-specific (Raykov, 1998; Curran & Bollen, 2001). Thus, LGC models are more important when the question of interest relates to individual-time paths or when one is interested in correlates of change over time (Raykov, 1998). When there is not significant growth over time or when individual differences are not of interest, autoregressive models may be more useful (MacKinnon, 2008a; Selig & Preacher, 2009). Curran and Bollen (2001,

p. 109) succinctly summarize the differences: “The Markov simplex modeling approach is well suited for examining the time-specific relations between two constructs over time and the growth modeling approach is well suited for examining relations in individual differences in continuous developmental trajectories over time.”

With respect to selecting an appropriate model for assessing longitudinal mediation, similar issues can be raised (i.e., is the interest in differences in individual trajectories and how such trajectories on a set of different variables relate to one another or is one interested in time-specific relations among the variables). Although the assessment of the mediated effect is more straightforward with LGC models, the autoregressive model is more conducive to examining complex patterns of mediation across waves and the possibility of cross-lagged relationships. Furthermore, if one is interested in whether relationships among X, M, and Y vary over time or in the assessment of when a mediational relationship starts or ends in a longitudinal study, the autoregressive model may have certain advantages (Cheong, 2011). Aside from these issues, it is difficult to compare these models. The indirect effects estimated by these two approaches, while both addressing the question of mediation, are in theory not the same population parameter. While it may be useful to compare the power and Type I errors associated with the detection of a mediated effect, standard methods for evaluating and comparing estimators are less relevant (e.g., degree of bias, consistency, efficiency).

MacKinnon (2008b) states that recommending one model over the other is difficult and suggests that in empirical practice it may be of use to estimate several longitudinal mediation models for the same data set. Convergence in evidence regarding the presence or absence of a mediated effect provides the most convincing evidence.

Maxwell and Cole (2007, p. 25) likewise “do not attempt to adjudicate which model is better” (i.e., the autoregressive mediation model or the random effects mediation model). Rather they consider the accuracy of estimating a longitudinal mediated effect with cross-sectional data first assuming the true process is an autoregressive model and then assuming the true process follows a random effects model. No direct comparisons are made. Similarly, papers two and three of this dissertation address issues specific to each of the two separate longitudinal mediation models, rather than directly comparing the two models. Paper two addresses issues with respect to the autoregressive mediation model and paper three addresses issues with respect to the LGC mediation model. Both of these papers implement simulation studies.

1.3 Latent Variables

Paper two directly incorporates latent variables, in this case, latent variables extracted from multiple indicators of the mediator at multiple time points. Paper three also utilizes latent variables, notably the growth parameters in a latent growth curve model. Thus, it may be useful to briefly discuss the concept of a latent variable.

Although there is some disagreement in the literature (e.g., see Bollen, 2002), perhaps the most straightforward definition of a latent variable is provided by Skrondal & Rabe-Hesketh (2004): “A *latent variable* [is] a random variable whose realizations are hidden from us. This is in contrast to *manifest variables* where the realizations are observed.” This distinction is probably most widely recognized in the social sciences, most notably in psychology, where latent variables are thought of as hypothetical constructs measured by multiple correlated observed (manifest) indicators (implemented

though the techniques of confirmatory factor analysis and structural equation modeling). Thus, self-esteem, personality, quality of life, intentions to perform a behavior, and mathematical ability are hypothetical constructs that are measured indirectly by using observed responses to a set of questionnaire items, for example. However, latent variables are used in a wide variety of applications and disciplines (although often with different names and terminology). Thus, economists refer to unobserved “measured-with-error” variables (the problem of errors-in-variables or EIV), which is equivalent to the concept of an unobserved latent variable modeled by psychologists and sociologists (Kennedy, 2003). Others from a variety of disciplines might use latent variable terminology such as common factors, random effects, unobserved heterogeneity, latent classes, latent traits (or theta from item response theory), counterfactuals, and latent growth curves (and still others).

Although a complete discussion of latent variables is beyond the scope of this paper, it is worth reviewing the process and benefits of extracting a latent variable from a set of manifest variables. Observed (or manifest) variables are linked to latent variables through a measurement model (sometimes called a confirmatory factor model) (Kaplan, 2009). Oftentimes, separate measurement models are specified for endogenous variables (Y-variables) and exogenous variables (X variables). These models are often presented using the notation from LISREL, a popular SEM program:

$$\begin{aligned} \text{Measurement model for } y: \mathbf{Y} &= \boldsymbol{\tau}_y + \boldsymbol{\Lambda}_y \boldsymbol{\eta} + \boldsymbol{\varepsilon} \\ \text{Measurement model for } x: \mathbf{X} &= \boldsymbol{\tau}_x + \boldsymbol{\Lambda}_x \boldsymbol{\xi} + \boldsymbol{\delta} \end{aligned} \tag{1.6}$$

The $\boldsymbol{\Lambda}$ matrices contain values known as factor loadings, which are essentially slope coefficients for the regression of the manifest variables, \mathbf{X} and \mathbf{Y} , on the latent variables, $\boldsymbol{\eta}$ and $\boldsymbol{\xi}$. The intercept terms, $\boldsymbol{\tau}$, do not really matter in confirmatory factor analysis

(CFA) because they do not contribute to covariance or affect model fit (unless a mean structure is tested) (i.e., researchers usually work with mean-centered variables). The relationships linking the exogenous and endogenous latent variables are specified in the structural model (note that some endogenous variables may serve as predictors of other endogenous variables):

$$\text{Structural model: } \boldsymbol{\eta} = \boldsymbol{\alpha} + \mathbf{B}\boldsymbol{\eta} + \boldsymbol{\Gamma}\boldsymbol{\xi} + \boldsymbol{\zeta} \quad (1.7)$$

The $\boldsymbol{\alpha}$ matrix contain intercepts, the \mathbf{B} matrix contains regression weights interrelating endogenous variables ($\boldsymbol{\eta}$), and the $\boldsymbol{\Gamma}$ matrix of regression weights relating exogenous variables ($\boldsymbol{\xi}$) to endogenous variables ($\boldsymbol{\eta}$), and zeta ($\boldsymbol{\zeta}$) is a vector of residuals for the endogenous latent variables.

The most critical benefit of extracting a latent variable from several manifest variables is that under certain assumptions (most notably that each set of manifest variables contains at least congeneric measures of the intended latent variable – Cole & Maxwell, 2003), the latent variables are without measurement error. Therefore, estimates of relationships among such variables are not biased by fallible (i.e., unreliable) measures (the effects of biased parameter estimates have subsequent effects on Type I and Type II errors, which may be minimized by the use of latent variables) (Bedeian, Day, & Kelloway, 1997). This is conceptually similar to the correction for attenuation formula based on the work of Spearman (1904) (Bedeian, Day, & Kelloway, 1997).

As mentioned earlier, parameter estimates in confirmatory factor analysis and SEM (both the measurement and structural models) are determined that yield a predicted variance-covariance matrix (one that is implied by the estimates of the unknowns) that

best reproduces the sample variance-covariance matrix. The traditional method of estimation in SEM is maximum likelihood, although many other options are available.

It is a straightforward conceptual extension to extract latent variables from a set of manifest variables in the autoregressive framework, although its implementation raises additional considerations that could affect the estimation of mediation effects. Many investigators still use linear composites of fallible indicators (e.g., sum scores or average scores of a set of indicators) to estimate underlying constructs, and while the negative impact of measurement errors can be minimized by using measures with higher levels of composite reliability, there are benefits of using latent variables even in this case. It is also possible to use a measurement model and extract latent variables from fallible indicators in a latent growth curve framework. Such models go by a variety of names including curve-of-factors models (McArdle, 1988; Duncan & Duncan, 2004; Liu et al., 2009) and second-order latent growth models (Hancock, Kuo, & Lawrence, 2001; Sayer & Cumsille, 2001).

1.4 Summary of Research Objectives

Below is a synopsis of the research objectives of each paper of the dissertation.

Paper one:

1. Assess whether the relationship between functional status (i.e., activities of daily living (ADLs) or what individuals perceive that they are able to do) and health-related quality of life is mediated by life-space mobility (i.e., what people actually do in terms of mobility) using an autoregressive longitudinal model of mediation.

Paper two:

1. Assuming an autoregressive longitudinal mediation model, assess the effects of the method to handle fallible indicators and the omission of paths representing shared method variance on parameter and standard error estimation as well as statistical power and Type I error rates under a variety of conditions (i.e., degree of shared method variance, degree of composite reliability for the set of indicators representing the mediator, degree of stability of the latent mediator, size of the mediated effect, and sample size).

Paper three:

1. Provide an overview of the two-stage piecewise parallel process latent growth curve model of mediation.
2. Evaluate the statistical performance (bias, power, and Type I error rate) of methods used to test mediation in this model under different conditions (i.e., degree of later growth in the mediator, degree of earlier growth in the outcome, complete or partial mediation, size of the mediated effect, and sample size).
3. Examine the impact of misspecifying the true piecewise model as a single-stage parallel process model of mediation under different conditions (i.e., degree of later growth in the mediator, degree of earlier growth in the outcome, complete or partial mediation, size of the mediated effect, and sample size).

Each of these objectives, including the motivation for them, will be discussed in more detail in the respective papers.

FUNCTIONAL STATUS, LIFE-SPACE MOBILITY, AND QUALITY OF LIFE:
A LONGITUDINAL MEDIATION ANALYSIS

by

BENTLEY, J. P., BROWN, C. J., MCGWIN, G., SAWYER, P. L., ALLMAN, R. M.,
AND ROTH, D. L.

In preparation for *Quality of Life Research*

Format adapted for dissertation

Abstract

Purpose. Using the Wilson-Cleary model of patient health outcomes as a conceptual framework, this study assessed whether the relationship between functional status (disability) and health-related quality of life is mediated by life-space mobility using data gathered from a population-based, longitudinal study of older adults.

Methods. Participants were enrollees in the University of Alabama at Birmingham (UAB) Study of Aging, a longitudinal study of mobility among community-dwelling older adults. Data were from four waves of the study spaced approximately 18 months apart from $n = 677$ participants who survived at least one year beyond the final assessment period. Both cross-sectional mediation models using baseline data and autoregressive mediation models using structural equation modeling were evaluated.

Results. The longitudinal analyses, based on an autoregressive model, supported the mediating role of life-space mobility and suggest that this role is more significant with the mental component summary score from the SF-12 as the outcome compared to the physical component summary score. Model modifications guided by theory and empirical findings did not alter the substantive meaning of this mediated effect, only enhanced it. The possibility of a reciprocal relationship over time between disability and life-space mobility was suggested by modification indices. Estimates of the mediation parameters from the autoregressive models were only partially consistent with mediation estimates from the cross-sectional analyses, suggesting that mediating relationships would have been missed or were potentially underestimated in the cross-sectional models.

Introduction

Mobility in older adults is often used as an outcome in health care as many interventions are designed to enhance the frequency and extent of individuals' movement as well as to minimize the amount of assistance needed. To this end, studies may examine the effects of interventions designed to enhance mobility or try to understand how biomedical, psychological, sociological, or environmental variables relate to, or predict, mobility (Baker, Bodner, & Allman, 2003; Peel et al., 2005). Mobility has also been shown to be related to the physical and mental components of quality of life (Baker et al., 2003), while others have suggested it might be useful to indicate global functional decline in older patients (Brown et al., 2009), suggesting the role of mobility as a useful predictor of other variables.

Self-reports of difficulty with activities of daily living (ADLs) are often used to assess functional ability (or disability). These are classic, clinically-relevant indicators that reflect what individuals perceive that they are able to do. In terms of affecting daily living, self-reported functional capability is just one variable that affects actual observed mobility, or what individuals actually do (Peel et al., 2005). The University of Alabama at Birmingham (UAB) Study of Aging Life-Space Assessment (LSA) was developed to measure individuals' patterns of mobility during the four weeks preceding an assessment (Baker et al., 2003). The instrument captures mobility in terms of extent and frequency of movement across a number of levels (i.e., with one's home up to outside one's town),

as well as the degree of independence in performing such movements. One's perception of the ability to carry out ADLs has been shown to be related to the LSA (Baker et al., 2003; Peel et al., 2005) and scores on the LSA have also been shown to be correlated to the physical and mental component summary scores of the SF-12 (Baker et al., 2003), a commonly used measure of health-related quality of life (Ware, Kosinski, & Keller, 2002). However, little effort has been directed at understanding the relationships among all three of these variables.

Wilson and Cleary (1995) developed a theoretical framework in an attempt to link the biomedical model of health with a quality of life model (Figure 1). Their model focuses on five levels of patient outcomes (i.e., biologic and physiologic variables, patient symptoms, patient functioning, overall or general health perceptions, and overall quality of life) and proposes mediating relationships among these variables (e.g., the relationship between symptom status and overall quality of life is mediated by functional status (or disability) and general health perceptions) in addition to explanatory roles of several personal and environmental variables. Parts of this model have been applied to a variety of patient populations and some authors have even expressly examined the mediated effects implied by the model (e.g., see Sousa & Kwok, 2006; Baker, Pankhurst, & Robinson, 2007; Ryu, West, & Sousa, 2009; Wyrwich, Harnam, Locklear, Svedsäter, & Revicki, 2011). With some adaptation, the Wilson-Cleary model provides a useful conceptual framework for examining the relationships among functional status, life space, and quality of life. The purpose of this study is to assess whether the relationship between functional status (i.e., ADLs or what individuals perceive that they are able to do) and health-related quality of life is mediated by life-space mobility (i.e., what people

actually do in terms of mobility) (see Figure 2 for conceptual view). Other studies have explored mediators of the relationship between functional status and quality of life. For example, using a sample of older adults, Newsom and Schulz (1996) showed that social support acts as a partial mediator between functional status and quality of life.

In addition to exploring this set of relationships, this study adds to the growing literature using longitudinal data to test mediation relationships (Cheong, MacKinnon, & Khoo, 2003; Cole & Maxwell, 2003; Maxwell & Cole, 2007; Little, Preacher, Selig, & Card, 2007; MacKinnon, 2008a; Liu et al., 2009; Audrain-McGovern, Rodriguez, & Kassel, 2009; Selig & Preacher, 2009; Negri, Ji, & Trickett, 2011; Roth & MacKinnon, 2012). In the case of true longitudinal mediation, Maxwell and Cole (2007) have demonstrated that cross-sectional analyses provide biased estimates of the indirect effect (or mediated effect) in two different models of change, the autoregressive model and a random effects model. Most previous analyses of mediated relationships based on the Wilson-Cleary model have relied on cross-sectional data (Sousa & Kwok, 2006; Baker et al., 2007; Ryu et al., 2009). Some evaluations of the Wilson-Cleary model have used longitudinal data, but did not test for mediated effects (e.g., Wilson & Cleary, 1997; Mathisen et al., 2007) or used only two waves of data (e.g., Wyrwich et al., 2011). Some advocate that three waves of data are necessary to truly test mediation (i.e., for a model to be a fully longitudinal mediation model) (Collins, Graham, & Flaherty, 1998). Although it is possible to provide some evidence of mediation with two waves of data, some additional assumptions must be made (Cole & Maxwell, 2003).

Methods

Setting and Participants

The participants in this study come from the UAB Study of Aging, a population-based, longitudinal study of mobility among community-dwelling older adults, sampled from a list of Medicare beneficiaries living in one of five counties (two urban and three rural) in central Alabama. The Study of Aging included $n = 1,000$ participants recruited between November 1999 and February 2001. African Americans, men, and rural residents were oversampled to provide a balanced sample in terms of race, gender, and rural/urban residence. Details of the study have been described previously (Peel et al., 2005). For the present analysis, data from four assessments (or waves) were analyzed. The waves were approximately 18 months apart (i.e., baseline, 18, 36, and 54 months after baseline). Participants who did not survive at least one year beyond the final assessment were excluded to limit the impact of end-of-life deterioration on the measures of interest. Of the 1,000 initial participants, $n = 250$ died prior to month 54 and $n = 61$ died after month 54 but before month 66. An additional $n = 12$ participants had no data other than baseline data and were excluded from analysis. Thus, $n = 677$ participants were used in the present analysis. As expected, compared to excluded individuals, participants in the present analysis were younger and at baseline had fewer verified comorbidities, had higher life space scores, had less difficulty in activities of daily living, and had higher health-related quality of life scores ($p < 0.0001$ for all comparisons). The original study, as well as the present analysis, was approved by the UAB Institutional Review Board and all participants gave written informed consent.

Measures

The primary measures of interest for this study are measures of functional disability (i.e., ADLs), life-space mobility, and health-related quality of life (i.e., the physical and mental component summary scores of the SF-12). All measures were self-report.

Disability. A latent disability measure based on categorical factor indicators was extracted using the following four ADL items treated as dichotomous observed variables, with 1 indicating the presence of difficulty with the task: transferring, bathing, dressing, and toileting. Mobility-related ADLs (i.e., using stairs, walking, and getting outside) were excluded due to their conceptual overlap with the life-space assessment. Data on two other ADLs were collected but excluded from the analysis because they were very infrequently endorsed by respondents (i.e., turning in bed, eating). At some waves, no respondents endorsed these items, prohibiting the extraction of a latent disability variable from these items at those waves.

Life-space mobility. The UAB LSA is described in more detail elsewhere (Baker et al., 2003; Peel et al., 2005). This instrument measures mobility based on the distance through which a person reports moving during the four weeks preceding the assessment. The composite measure of life-space (LS-C) was used as an observed variable in the present analysis. Scores for this variable ranged from 0 (“totally bedroom bound”) to 120 (traveled out of town every day without assistance). The two-week follow-up test-retest reliability of the LS-C measure was estimated to be 0.96 using the intraclass correlation coefficient (Baker et al., 2003). Given its high degree of reliability and its calculation as

a composite variable, LS-C was treated as an observed variable rather than a latent variable.

Health-related quality of life. Version 1 of the SF-12 was used to assess health-related quality of life. Mental component summary scores (MCS-12) and physical component summary scores (PCS-12) were calculated using procedures outlined by Ware et al. (2002). PCS-12 and MCS-12 are transformed such that the norm population (the general U.S. population) has a mean of 50 and a standard deviation of 10. Test-retest reliabilities of 0.89 and 0.76 have been reported for the PCS-12 and MCS-12, respectively (Ware et al., 2002). Given established, strong measurement properties and frequent use within the field, MCS-12 and PCS-12 were treated as observed variables in all analysis (i.e., latent variables were not extracted from the individual SF-12 items). Not only is this consistent with common practice, but prior research has shown very high correlations with these summary scores and latent variables based on individual items (0.97 for PCS and 0.96 for MCS) (Okonkwo, Roth, Pulley, & Howard, 2010).

Covariates. The following variables, all collected at baseline, were used as covariates in all analyses: sex, race, age, recent smoking status, education, and a verified co-morbidity score.

Statistical Analysis

Statistical mediation suggests that the causal influence of one variable on another is transmitted through a third variable called a mediator or intervening variable (Hoyle and Kenny, 1999). A variety of statistical methods for assessing mediation have been proposed (e.g., see MacKinnon, 2008b). The presence of longitudinal data provides

additional options for data analysis (Cole & Maxwell, 2003; Maxwell & Cole, 2007; MacKinnon, 2008a, Roth & MacKinnon, 2012). It is often difficult to recommend one approach over another. The present analysis evaluated the proposed mediated effect utilizing an autoregressive model based on a structural equation modeling (SEM) approach.

The basic premise of autoregressive models is that the values of a variable at a future time point depend in some part on an earlier time point. The simplest of these models involves one variable measured at multiple time points for a set of individuals. This basic model (the univariate model) is often referred to the simplex model (or the Markov simplex model) (Marsh, 1993; Curran & Bollen, 2001) and dates back to the work of Guttman (1954). Gollob and Reichardt (1991) and subsequently Cole and Maxwell (2003) extended the univariate simplex model to a trivariate model, an autoregressive mediation model, involving the variables: X (the independent variable), M (the mediator), and Y (the outcome). The autoregressive mediation model is preferred over other options (e.g., latent growth curve (LGC) mediation models or parallel process models) when the variables of interest do not show evidence of significant change over time (i.e., the autoregressive mediation model is considered a model of interindividual change rather than intraindividual change) (MacKinnon, 2008a; Selig & Preacher, 2009). In the present case, some of the variables did not exhibit considerable change over the assessment periods suggesting that an autoregressive mediation model is appropriate. For example, even though there was a statistically significant and linear change over time in MCS-12, the scores were relatively stable during the study period with an average change over each 18-month observation period of less than 0.05 of a standard deviation.

Other tests of the Wilson-Cleary model have used autoregressive modeling. For example, Mathisen and colleagues (2007) used four waves of data to test the link between general health perceptions and overall quality of life, a relationship implied by the Wilson-Cleary model. Although they did not test mediation, they did use SEM to examine cross-lagged effects and simultaneous effects in the context of an autoregressive model, addressing the question of whether reciprocal effects are possible in the Wilson-Cleary model.

A basic framework for the analysis can be found in Figure 3. For the sake of clarity, only structural paths (i.e., no measurement paths, no residual covariances) for the main variables (i.e., no adjustments for covariates) are included in Figure 3. Separate models were fit for MCS-12 and PCS-12 as outcome variables.

A modified set of steps outlined by Cole and Maxwell (2003) as well as MacKinnon (2008a) for autoregressive mediation modeling was utilized. Prior to fitting any longitudinal models, a cross-sectional assessment of mediation was evaluated using baseline data. Next, an initial autoregressive mediation model (i.e., Model 1's) was separately fit for MCS-12 and PCS-12 as outcome variables based on the paths outlined in Figure 3. Although measurement invariance over time for the disability measure based on ADLs was not assumed (i.e., factor loadings and thresholds were free to vary over time), these models did include covariates as control variables as well as covariances between residual terms of downstream endogenous variables within the same wave. This last model component is recommended by Anderson and Williams (1992) and the presence of significant and meaningful covariation among these residuals implies the possibility of confounding due to variables missing from the model and failing to specify

such covariances may lead to biased estimates (Cole & Maxwell, 2003). Following this step, a test of measurement invariance over time for the disability measure was performed. The indirect effect (or mediated effect) of disability on health-related quality of life through life-space mobility was then calculated and evaluated for statistical significance using both the Sobel test (Sobel, 1982) and a bias-corrected bootstrap confidence interval with 1,000 draws (MacKinnon, Lockwood, & Williams, 2004). With four waves of data, it is possible to estimate both time-specific indirect effects, as well as the overall (or total) indirect effect (Gollob & Reichardt, 1991; Cole & Maxwell, 2003; Selig & Preacher, 2009). As outlined in Figure 3, there are three time-specific indirect effects (with the estimate of each indirect effect based on the product of coefficients method):

1. Disability₁ → Disability₂ → LS₃ → PCS₄ (or MCS₄) = $x_1 * a_2 * b_3$
2. Disability₁ → LS₂ → LS₃ → PCS₄ (or MCS₄) = $a_1 * m_2 * b_3$
3. Disability₁ → LS₂ → PCS₃ (or MCS₃) → PCS₄ (or MCS₄) = $a_1 * b_2 * y_3$

The overall indirect effect of Disability₁ → PCS₄ (or MCS₄) is the sum of the three time-specific indirect effects. This addresses the question of whether life-space mobility mediates the effect of disability₁ on PCS₄ (or MCS₄) at any time between waves 1 and 4 (rather than at some specific point), a question usually of greater interest to researchers (Cole & Maxwell, 2003). Indeed, Selig and Preacher (2009, p. 150) note that the “sum of all possible indirect effects more faithfully depicts the degree to which X₁ indirectly influences Y₄.” Thus, both time-specific and overall indirect effects were calculated and evaluated for significance.

In addition to the measure of indirect effect, the fit of the models to the data was evaluated using various indices, including the chi-square goodness of fit statistic, the root

mean square error of approximation (RMSEA), comparative fit index (CFI), and the weighted root mean square residual (WRMR). Good model fit is indicated by a small chi-square value (this test is quite sensitive to sample size), an RMSEA less than 0.05 (Brown & Cudeck, 1993), a CFI above 0.95 (Hu & Bentler, 1999), and a WRMR of less than 1.0 (Yu, 2002). The PCLOSE index provides a probability that the RMSEA for a given model is less than or equal to 0.05.

Following the test of the indirect effect, the potential for omitted paths was evaluated using modification indices. Model 2's based on the results of these specification searches, were then estimated separately for MCS-12 and PCS-12 as outcome variables. In essence, this step allowed for the inclusion of fit-based model enhancements, some of which might have the potential to alter the size and/or significance of the indirect effect of interest. Thus, following this step, tests of the mediated effects were performed again. Potential fit-based model adjustments that were evaluated included modifications that could be justified theoretically such as: 1) covariances between measurement errors for the same ADLs measured at several time points, 2) wave-skipping autoregressive paths (e.g., disability₄ predicted by disability₃ and disability₂), 3) cross-lagged relations (e.g., disability₂ predicted by LS₁), 4) autocorrelation residuals (e.g., residual variance of disability₃ with the residual variance of disability₂), 5) contemporaneous paths (e.g., LS₂ predicted by disability₂), and additional direct effects (e.g., PCS₄ (or MCS₄) predicted by disability₂). Effects going backward in time (e.g., time₁ variable predicted by a time₂ variable) were not considered.

All analyses were conducted using SEM in *Mplus* 6.11 (Muthén & Muthén, Los Angeles, CA). To account for the categorical disability indicators (i.e., ADLs), a robust

weighted least squares estimator (i.e., WLSMV) was used (Muthén, 1984). Although all participants provided wave 1 data (including covariates), as with most longitudinal studies, data for some participants were not available for the remaining three waves. There were $n = 677$ complete cases at wave 1, $n = 651$ at wave 2, $n = 626$ at wave 3, and $n = 611$ at wave 4. In some cases, participants supplied life-space mobility and ADL scores, but not SF-12 scores. Rather than listwise deletion (i.e., analyzing only those with complete data at all four waves), the WLSMV estimator in *Mplus* uses the total available sample for analysis. Such estimates are consistent under the assumption of missing at random with respect to covariates (MARX) and much more efficient than listwise deletion (Asparouhov & Muthén, 2010).

Results

Table 1 provides means and standard deviations or frequencies for the main study variables as well as the variables used as covariates (baseline values). In addition, the numbers of available cases for each variable at each time point are provided.

Cross-Sectional Mediation Analysis

As a basis for comparison, prior to fitting the autoregressive mediation models, the mediating role of life-space mobility in the relationship between functional status and health-related quality of life was evaluated using baseline values. This cross-sectional analysis was conducted on the $n = 677$ cases used in the longitudinal analysis as well as all cases available at baseline ($n = 1000$). The latter analysis was conducted as it is likely that this is the approach that would have been used to evaluate the mediated relationship

had longitudinal data not been available. Using the terminology in Figure 2, the indirect effect was estimated by the product of coefficients method (ab). The proportion mediated effect was calculated as $ab/(c'+ab)$ (MacKinnon, 2008b).

The results of the cross-sectional models can be found in Table 2. In all cases, model fit was acceptable. With PCS-12 as the outcome variable, the mediated effect was statistically significant using both the Sobel test and the bias-corrected bootstrap CI method for the participants used in the longitudinal analysis ($n = 677$), but not when data from all participants ($n = 1000$) were analyzed. Given the relatively large direct effect of disability on PCS-12 ($c' = -8.20$, $SE=0.802$, $p<0.0001$), only approximately 5% of the total effect of disability on PCS-12 is mediated through life-space mobility. With MCS-12 as the outcome variable, the mediated effect was statistically significant for participants used in the longitudinal analysis ($n = 677$) and for all participants ($n = 1000$) using both the Sobel test and the bias-corrected bootstrap CI method. About 21% of the total effect of disability on MCS-12 is mediated by life-space mobility.

Longitudinal Mediation Analysis

Because disability was conceptualized as a latent variable represented by multiple measures (ADLs), it was possible to assess factorial invariance across waves. To assess invariance, initial models based on Figure 3 (with separate models for PCS-12 and MCS-12 as outcome variables) were estimated allowing factor loadings and thresholds to freely vary over time and compared to models where thresholds and loadings were constrained to equal their counterparts at subsequent waves. The WLSMV estimator in *Mplus* precludes the use of traditional chi-square difference testing to evaluate invariance. Thus,

the DIFFTEST option in *Mplus* was used (Muthén & Muthén, 1998-2010). As with traditional chi-square difference testing, a significant finding indicates a lack of measurement invariance. For both the PCS-12 and MCS-12 models, this test for the lack of invariance was statistically significant ($\Delta\chi^2 = 37.65$, $df = 21$, $p = 0.0142$; $\Delta\chi^2 = 42.21$, $df = 21$, $p = 0.004$ for PCS-12 and MCS-12 models, respectively). However, for relatively large samples, the test for invariance can be statistically significant (i.e., indicating a lack of invariance) even when there are only minor differences in model parameters (Millsap, 2005; Kline, 2011). Examination of fit indices that take model parsimony into account can provide information as to whether this is the case. The value of the RMSEA was actually reduced from the unconstrained to the invariant models ($0.041 \rightarrow 0.038$ and $0.039 \rightarrow 0.037$, for PCS-12 and MCS-12 models, respectively) and the value of CFI was either increased or remained unchanged ($0.981 \rightarrow 0.982$ and $0.983 \rightarrow 0.983$, for PCS-12 and MCS-12 models, respectively) with the addition of invariance constraints. This provides evidence that the deviations from perfect measurement invariance suggested by the statistical tests were small and of no practical significance. Thus, all subsequent models were tested with the assumption of invariance over time for the disability measure.

The results of the longitudinal mediation models with measurement invariance constraints imposed for disability can be found in Table 3. Model 1's for PCS-12 and MCS-12 as the outcome variables are based on the basic framework in Figure 3 and in both cases, the models fit the data well. With PCS-12 as the outcome variable, only one of the time-specific mediated effects was statistically significant, but the total indirect effect was statistically significant using both the Sobel test and the bias-corrected

bootstrap CI method (95% CI: -1.282, -0.062). About 13% of the total effect of disability on PCS-12 is mediated through life-space mobility. With MCS-12 as the outcome variable, all three time-specific mediated effects were statistically significant as well as the total indirect effect (using both the Sobel test and the bias-corrected bootstrap CI method (95% CI: -1.17, -0.235)). About 44% of the total effect of disability on MCS-12 is mediated by life-space mobility. However, the direct effect of disability on MCS-12 was only marginally significant ($p = 0.051$), suggesting the possibility of complete mediation of the disability and MCS-12 relationship by life-space mobility.

Modification indices were used as a guide to assess the potential for omitted paths. Final Model 2's with the additional paths can be found in Table 3. In addition, completely standardized path coefficients for the original paths as well as the newly added paths can be found in Figures 4 and 5 for the PCS-12 and MCS-12 models, respectively. The newly added paths consisted of either wave-skipping autoregressive paths (e.g., disability₄ predicted by disability₂) or cross-lagged relations (e.g., disability₂ predicted by LS₁). Modification indices suggested no other paths that would lead to substantial improvements in model fit after these paths were included. Thus, no additional paths were added for covariances between measurement errors for the same ADLs measured at several time points, autocorrelation residuals, contemporaneous paths, or additional direct effects. As expected, model fit improved significantly with the addition of the paths listed in Table 3. More importantly, the significance of the disability → life-space mobility → HRQOL mediated effect was maintained. According to the Model 2's, almost 20% of the total effect of disability on PCS-12 is mediated through life-space mobility, while it appears that all of the relationship between disability

and MCS-12 is mediated by life-space mobility (i.e., $p = 0.691$ for the direct effect of disability on MCS-12).

Discussion

In this study, a longitudinal analysis supported the existence of the mediating role of life-space mobility in the relationship between functional status (disability) and health-related quality of life, a relationship adapted from the Wilson-Cleary model of patient health outcomes. Moreover, estimates of the mediational parameters in the autoregressive models were only partially consistent with estimates cross-sectional analyses. For example, researchers might have concluded that there was no significant mediation of the disability and PCS-12 relationship using all available cases at baseline. In all cases, the longitudinal models revealed a larger indirect effect as a percentage of the total effect when compared to cross-sectional models, even suggesting complete mediation when MCS-12 was the outcome variable.

Few rigorous analyses of mediation mechanisms using longitudinal designs have been published. Furthermore, to our knowledge, this is the first test of mediated relationships implied by the Wilson-Cleary model to use longitudinal data with three or more waves of data. There are several benefits of using longitudinal rather than cross-sectional data to evaluate mediation processes (MacKinnon, 2008a; Roth & MacKinnon, 2012). First, more information regarding the temporal sequentiality of the independent variable, the mediator, and the outcome variable is provided with longitudinal data, a critical underlying assumption of mediation. Second, longitudinal data allow for an examination of associations within waves of data (i.e., cross-sectional, between subjects) and changes across waves of data (i.e., within individuals). Associations between

variables in cross-sectional models capture both time-varying and time-invariant covariances between participants, but longitudinal models allow these two sources of covariance to be separated and tested more directly in line with the causal assumptions of the mediation model. Finally, individuals may serve as their own control in assessing some relationships, thereby potentially controlling for static differences among individuals. In their discussion of why cross-sectional data generally provide poor estimates of effects, Gollob and Reichardt (1991) note two other related benefits to the use of longitudinal data to assess mediation. First, causal effects often take time to develop, and variables measured at the same time may not allow for the necessary development time. Second, variables often have effects on themselves, such that an outcome variable at a later time is related to the same outcome variable at an earlier time. Cross-sectional models implicitly assume these autoregressive effects are zero, and substantial bias can be introduced as a result. Cole and Maxwell (2003) note that it is not sufficient to merely allow a time lag between the independent variable and the mediator and then between the mediator and the outcome to achieve unbiased estimates of effects because of the potentially confounding effects of prior levels of the mediator and the outcome. Autoregressive effects control for these relationships, and accordingly adjust the time-varying associations that are being tested as part of the causal model.

The mediating role of life-space mobility appears to be more significant with the mental component summary score from the SF-12 as the outcome compared to the physical component summary score. The final level of the Wilson-Cleary model, overall quality of life, has been represented differently by different authors. Some have used health-related quality of life measures (generic or disease-specific) (e.g., Höfer et al.,

2005; Heo et al., 2005; Ryu et al., 2009) while others have used life satisfaction measures (e.g., Mathisen et al., 2007; Wyrwich et al., 2011) or measures of mental health (Sousa & Kwok, 2006) and psychological distress (Baker et al., 2007). The items in the SF-12 most strongly related to PCS-12 are more closely related to physical functioning. Höfer and colleagues used the physical functioning scale of the SF-36, which is a large part of the PCS scores, as a measure of functional status. Given the conceptual overlap between disability, life-space mobility, and PCS-12, it is not surprising that the mediating role of life-space mobility is less significant with PCS-12 as an outcome relative to MCS-12. The MCS-12 seems more consistent with the end level in the Wilson-Cleary model.

Although not tested as an a priori hypothesis, the existence of cross-lagged relationships between disability and life space mobility as indicated by the modification indices, suggest that disability levels not only influence changes in life-space mobility, but also show effects in response to changes in mobility. Other evaluations of the Wilson-Cleary model have examined such reciprocal effects. For example, Mathisen and colleagues (2007) note reciprocal causal effects over time between general health perceptions and overall quality of life, the fourth and fifth levels in the Wilson-Cleary model. Wilson and Cleary (1995) note the possibility of such effects and other researchers have called for an examination of these bidirectional relationships (Baker et al., 2007; Wyrwich et al., 2011). Further research is required to support the view that the Wilson-Cleary model be modified to explicitly incorporate bidirectional causal effects.

There are several limitations of the present study that should be considered. Although the longitudinal analysis conducted in this study has several benefits over cross-sectional views of mediation, it is important to recognize that the data analyzed in

the present study are still considered observational. Longitudinal structural equation modeling cannot prove causality in these kinds of designs, but it can test whether a proposed causal model is plausible and whether it provides good fit to the observed data. In the present study, 18 months separated each wave. The timing and spacing of measurements can influence the detection and estimation of effects in longitudinal studies (Collins & Graham, 2002), and the 18-month interval may not have been optimal to detect more rapidly transmitting causal effects. Missing data is usually problematic in any longitudinal study. Given the study that generated the data was of older adults, it is not surprising that a substantial number of participants did not survive the entire study period. These individuals were excluded as the intent of the study was to examine the mediated effect of interest in a population not experiencing the rapid decline in mobility and health that often corresponds with an end-of-life deterioration. A smaller set of individuals failed to complete certain assessments during the study period. In the analysis of the data, we assumed that the missing data mechanism in this data set was missing at random with respect to covariates (MARX) (Asparouhov & Muthén, 2010). For other missing data mechanisms, other approaches such as multiple imputation (Schafer, 1997) or pattern mixture modeling (Hedeker & Gibbons, 1997) would be more appropriate. Finally, although the model modification through specification searches only considered theoretical plausible paths, it is important to note that such modifications can lead to model misspecifications.

Conclusion

A mediation relationship implied by the Wilson-Cleary model of patient health outcomes was evaluated. Cross-sectional models may be inadequate to fully explain the role of life-space mobility in mediating the relationship between disability and quality of life. The present longitudinal analysis, based on an autoregressive model, supports the mediating role of life-space mobility and suggests that this role is more significant with the mental component summary score from the SF-12 as the outcome compared to the physical component summary score. Furthermore, model modifications guided by theory and empirical findings did not alter the substantive meaning of this mediated effect, only enhanced it. In addition, the possibility of reciprocal relationships, consistent with other studies of the Wilson-Cleary model, is suggested. Several authors have advocated for the use of longitudinal designs in evaluating relationships proposed in the Wilson-Cleary model. The present study provides an example of one approach for testing mediation with longitudinal data.

References

- Anderson, S. E., & Williams, L. J. (1992). Assumptions about unmeasured variables with studies of reciprocal relationships: The case of employee attitudes. *Journal of Applied Psychology*, 77, 638-650.
- Asparouhov, T., & Muthén, B. (2010). Weighted least squares estimation with missing data. Retrieved from <http://www.statmodel.com/download/GstrucMissingRevision.pdf>.
- Audrain-McGovern, J., Rodriguez, D., & Kassel, J. D. (2009). Adolescent smoking and depression: Evidence for self-medication and peer smoking mediation. *Addiction*, 104, 1743-1756.
- Baker, P. S., Bodner, E. V., & Allman, R. M. (2003). Measuring life-space mobility in community-dwelling older adults. *Journal of the American Geriatrics Society*, 51, 1610-1614.

- Baker, S. R., Pankhurst, C. L., & Robinson, P. G. (2007). Testing relationships between clinical and non-clinical variables in xerostomia: A structural equation model of oral health-related quality of life. *Quality of Life Research*, 16, 297-308.
- Brown, C. J., Roth, D. L., Allman, R. M., Sawyer, P., Ritchie, C. S., & Roseman, J. M. (2009). Trajectories of life-space mobility after hospitalization. *Annals of Internal Medicine*, 150, 372-378.
- Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 136-162). Newbury Park, CA: SAGE.
- Charlson, M., Pompei, P., Ales, K., & MacKenzie, C. (1987). A new method of classifying prognostic comorbidity in longitudinal studies: Development and validation. *Journal of Chronic Diseases*, 40, 373-383.
- 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.
- Cole, D. A., & Maxwell, S. E. (2003). Testing mediation 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.
- Collins, L. M., & Graham, J. W. (2002). The effect of the timing and spacing of observations in longitudinal studies of tobacco and other drug use: temporal design considerations. *Drug and Alcohol Dependence*, 68, S85-S96.
- Curran, P. J., & Bollen, K. A. (2001). The best of both worlds: Combining autoregressive and latent curve models. In L. M. Collins & A. G. Sayer (Eds.), *New methods for the analysis of change* (pp. 107-135). Washington, DC: American Psychological Association.
- 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.
- Guttman, L. A. (1954). A new approach to factor analysis: The radex. In P. F. Lazarsfeld (Ed.), *Mathematical thinking in the social sciences* (pp. 258-348). New York: Columbia University Press.
- Hedeker, D. & Gibbons, R. D. (1997). Application of random-effects pattern-mixture models for missing data in longitudinal studies. *Psychological Methods*, 2, 64-78.
- Heo, S., Moser, D., Riegel, B., Hall, L., & Christman, N. (2005). Testing a published model of health-related quality of life in heart failure. *Journal of Cardiac Failure*, 11, 372-379.
- Höfer, S., Benzer, W., Alber, H., Ruttman, E., Kopp, M., Schüssler, G., & Doering, S. (2005). Determinants of health-related quality of life in coronary artery disease patients: A prospective study generating a structural equation model. *Psychosomatics*, 46, 212-223.

- Hoyle, R. H., & Kenny, D. A. (1999). Sample size, reliability, and tests of statistical mediation. In R. H. Hoyle (Ed.), *Statistical strategies for small sample research* (pp. 195-222). Thousand Oaks, CA: Sage.
- Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6, 1-55.
- Kline, R. B. (2011). *Principles and practice of structural equation modeling*, 3rd ed. New York: The Guilford Press.
- Little, T. D., Preacher, K. J., Selig, J. P., & Card, N. A. (2007). New developments in latent variable panel analyses of longitudinal data. *International Journal of Behavioral Development*, 31, 357-365.
- Liu, L. C., Flay, B. R., & Aban Aya Investigators. (2009). Evaluating mediation in longitudinal multivariate data: Mediation effects for the Aban Aya youth project drug prevention program. *Prevention Science*, 10, 197-207.
- 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. (2008a). Longitudinal mediation analysis. In D.P. MacKinnon, *Introduction to statistical mediation analysis* (pp. 193-236). New York: Lawrence Erlbaum Associates.
- MacKinnon, D. P. (2008b). *Introduction to statistical mediation analysis*. New York: Lawrence Erlbaum Associates.
- Mathisen, L., Andersen, M. H., Veenstra, M., Wahl, A. K., Hanestad, B. R., & Fosse, E. (2007). Quality of life can both influence and be an outcome of general health perceptions after heart surgery. *Health and Quality of Life Outcomes*, 5, 27.
- Maxwell, S. E., & Cole, D. A. (2007). Bias in cross-sectional analyses of longitudinal mediation. *Psychological Methods*, 12, 23-44.
- Millsap, R. E. (2005). Four unresolved problems in studies of factorial invariance. In A. Maydeu-Olivares & J. J. McArdle (Eds.), *Contemporary psychometrics: A festschrift to Roderick P. McDonald* (pp. 153-171). Mahwah, NJ: Erlbaum.
- Muthén, B. (1984). A general structural equation model with dichotomous, ordered categorical, and continuous latent variable indicators. *Psychometrika*, 49, 115-132.
- Muthén, L. K. and Muthén, B. O. (1998-2010). *Mplus User's Guide*. Sixth Edition. Los Angeles, CA: Muthén & Muthén.
- Negriff, S., Ji, J., & Trickett, P. (2011). Exposure to peer delinquency as a mediator between self-report pubertal timing and delinquency: a longitudinal study of mediation. *Development and Psychopathology*, 23, 293-304.
- Newsom, J. T., & Schulz, R. (1996). Social support as a mediator in the relation between functional status and quality of life in older adults. *Psychology and Aging*, 11, 34-44.
- Okonkwo, O. C., Roth, D. L., Pulley, L., & Howard, G. (2010). Confirmatory factor analysis of the validity of the SF-12 for persons with and without a history of stroke. *Quality of Life Research*, 19, 1323-1331.

- Peel, C., Baker, P. S., Roth, D. L., Brown, C. J., Bodner, E. V., & Allman, R. M. (2005). Assessing mobility in older adults: The UAB Study of Aging Life-Space Assessment. *Physical Therapy*, 85, 1008-1019.
- Roth, D. L., & MacKinnon, D. P. (2012). Mediation analysis with longitudinal data. In J. T. Newsom, R. N. Jones, & S. M. Hofer (Eds.), *Longitudinal data analysis: A practical guide for researchers in aging, health, and social sciences* (pp. 181-216). New York: Routledge.
- Ryu, E., West, S. G., & Sousa, K. H. (2007). Mediation and moderation: Testing relationships between symptom status, functional health, and quality of life in HIV patients. *Multivariate Behavioral Research*, 44, 213-232.
- Schafer, J. L. (1997). *Analysis of incomplete multivariate data*. London: Chapman & Hall.
- Selig, J. P., & Preacher, K. J. (2009). Mediation models for longitudinal data in developmental research. *Research in Human Development*, 6, 144-164.
- Sobel, M. E. (1982). Asymptotic confidence intervals for indirect effects in structural equation models. *Sociological Methodology*, 13, 290-312.
- Sousa, K., & Kwok, O. (2006). Putting Wilson and Cleary to the test: Analysis of a HRQOL conceptual model using structural equation modeling. *Quality of Life Research*, 15, 725-737.
- Ware, J. E., Kosinski, M., & Keller, S.D. (2002). *SF-12: How to Score the SF-12 Physical and Mental Health Summary Scales*, 4th ed. Lincoln, RI: QualityMetric Incorporated.
- Wilson, I. B., & Cleary, P. D. (1995). Linking clinical variables with health-related quality of life: A conceptual model of patient outcomes. *JAMA*, 272, 59-65.
- Wilson, I., & Cleary, P. (1997). Clinical predictors of declines in physical functioning in persons with AIDS: Results of a longitudinal study. *Journal of Acquired Immune Deficiency Syndromes and Human Retrovirology*, 16, 343-349.
- Wyrwich, K., Harnam, N., Locklear, J., Svedsäter, H., & Revicki, D. (2011). Understanding the relationships between health outcomes in generalized anxiety disorder clinical trials. *Quality of Life Research*, 20, 255-262.
- Yu, C. Y. (2002). Evaluating cutoff criteria of model fit indices for latent variable models with binary and continuous outcomes. Doctoral dissertation, University of California, Los Angeles.

Table 1 Descriptive statistics

	Means [or <i>n</i> (%)]	SD
Covariates at baseline (<i>n</i> = 677)		
Sex – male	306 (45.2%)	
Race – African American	337 (49.8%)	
Recent smoker ¹	83 (12.3%)	
Age	74.09	6.02
Education ²	2.63	1.11
Comorbidity score ³	1.96	1.41
Wave 1 (<i>n</i> = 677)		
ADLs		
Transferring	177 (26.1%)	
Bathing	81 (12.0%)	
Dressing	64 (9.5%)	
Toileting	31 (4.6%)	
Life Space	69.12	22.94
PCS-12	42.19	12.63
MCS-12	54.33	9.03
Wave 2		
ADLs (<i>n</i> = 663)		
Transferring	114 (17.2%)	
Bathing	83 (12.5%)	
Dressing	73 (11.0%)	
Toileting	15 (2.3%)	
Life Space (<i>n</i> = 664)	64.74	22.38
PCS-12 (<i>n</i> = 651)	43.74	10.34
MCS-12 (<i>n</i> = 651)	54.95	7.38
Wave 3		
ADLs (<i>n</i> = 645)		
Transferring	74 (11.5%)	
Bathing	88 (13.6%)	
Dressing	56 (8.7%)	
Toileting	26 (4.0%)	
Life Space (<i>n</i> = 646)	62.77	23.58
PCS-12 (<i>n</i> = 627)	44.34	11.21
MCS-12 (<i>n</i> = 627)	55.41	7.17
Wave 4		
ADLs (<i>n</i> = 638)		
Transferring	124 (19.4%)	
Bathing	112 (17.5%)	
Dressing	71 (11.1%)	
Toileting	87 (13.6%)	
Life Space (<i>n</i> = 640)	57.88	25.92
PCS-12 (<i>n</i> = 612)	41.66	12.23
MCS-12 (<i>n</i> = 612)	55.14	7.60

Notes: ¹Recent smoker status includes those who are current smokers and those who quite within the past 12 months. ²Education is divided into four categories: 1 ≤ 6; 2 = 7-11; 3=13; 4 ≥ 13. ³Comorbidity score was created by summing the total number of verified comorbidities that make up the Charlson Comorbidity Index (Charlson Pompei, Ales, & MacKenzie, 1987).

Table 2 Model fit indices and unstandardized path coefficients for cross-sectional mediation models

	PCS as outcome <i>n</i> = 677	PCS as outcome <i>n</i> = 1000	MCS as outcome <i>n</i> = 677	MCS as outcome <i>n</i> = 1000
Path	Estimate (SE)	Estimate (SE)	Estimate (SE)	Estimate (SE)
Disability → LS	-8.74 (1.009)**	-13.29 (0.925)**	-9.01 (1.043)**	-13.67 (0.963)**
LS → PCS/MCS	0.049 (0.024)*	0.033 (0.02)	0.050 (0.020)*	0.057 (0.016)**
Disability → PCS/MCS	-8.20 (0.802)**	-9.07 (0.753)**	-1.67 (0.610)**	-2.98 (0.552)**
Indirect effect				
Disability→LS→PCS/MCS (ab)	-0.433 (0.204)*	-0.439 (0.260)*	-0.448 (0.182)*	-0.773 (0.216)**
<i>p</i> – value (Sobel test)	<i>p</i> = 0.034	<i>p</i> = 0.091	<i>p</i> = 0.014	<i>p</i> < 0.0001
95% bias-corrected bootstrap CI	(-0.932, -0.027)	(-0.986, 0.150)	(-0.937, -0.121)	(-1.295, -0.269)
χ^2 / df	44.81/26	74.72/26	37.81/26	64.81/26
RMSEA	0.033	0.043	0.026	0.039
PCLOSE	0.965	0.820	0.993	0.941
CFI	0.989	0.985	0.992	0.986
WRMR	0.702	0.870	0.643	0.813

Notes: Measurement paths (i.e., for disability) and covariate paths have been omitted from the table for simplicity but were included in each model. RMSEA = root mean square error of approximation, PCLOSE = probability that RMSEA ≤ 0.05, CFI = comparative fit index, WRMR = weighted root mean square residual. Standard errors (SE) for the indirect effect are from Sobel (1982).

* $p \leq 0.05$; ** $p \leq 0.01$

Table 3 Model fit indices and unstandardized path coefficients for longitudinal mediation models

Path	PCS as outcome Model 1	PCS as outcome Model 2	MCS as outcome Model 1	MCS as outcome Model 2
	Estimate (SE)	Estimate (SE)	Estimate (SE)	Estimate (SE)
Disability ₁ → Disability ₂	0.88 (0.046)**	0.65 (0.060)**	0.88 (0.047)**	0.62 (0.066)**
Disability ₂ → Disability ₃	0.84 (0.030)**	0.52 (0.068)**	0.85 (0.031)**	0.51 (0.071)**
Disability ₃ → Disability ₄	0.86 (0.030)**	0.65 (0.064)**	0.85 (0.030)**	0.64 (0.066)**
LS ₁ → LS ₂	0.21 (0.045)**	0.33 (0.042)**	0.29 (0.064)**	0.31 (0.044)**
LS ₂ → LS ₃	0.30 (0.061)**	0.40 (0.061)**	0.35 (0.066)**	0.28 (0.055)**
LS ₃ → LS ₄	0.05 (0.074)	-0.04 (0.075)	0.10 (0.074)	0.01 (0.091)
PCS/MCS ₁ → PCS/MCS ₂	0.60 (0.054)**	0.58 (0.056)**	0.29 (0.024)**	0.29 (0.024)**
PCS/MCS ₂ → PCS/MCS ₃	0.76 (0.061)**	0.31 (0.060)**	0.43 (0.032)**	0.40 (0.031)**
PCS/MCS ₃ → PCS/MCS ₄	0.53 (0.052)**	0.51 (0.056)**	0.43 (0.036)**	0.33 (0.031)**
Disability ₁ → LS ₂	-8.53 (1.353)**	-7.27 (1.043)**	-6.73 (1.844)**	-7.06 (1.146)**
Disability ₂ → LS ₃	-7.84 (1.330)**	-6.26 (1.212)**	-6.99 (1.404)**	-5.42 (1.251)**
Disability ₃ → LS ₄	-13.98 (1.719)**	-15.47 (1.723)**	-12.92 (1.682)**	-14.43 (2.001)**
LS ₁ → PCS/MCS ₂	0.03 (0.016)*	-0.004 (0.021)	0.07 (0.016)**	0.08 (0.016)**
LS ₂ → PCS/MCS ₃	0.09 (0.021)**	0.13 (0.021)**	0.06 (0.016)**	0.07 (0.017)**
LS ₃ → PCS/MCS ₄	0.02 (0.026)	0.06 (0.026)*	0.05 (0.017)**	0.09 (0.018)**
Disability ₁ → PCS/MCS ₄	-4.07 (0.666)**	-3.59 (0.693)**	-0.77 (0.392)†	0.16 (0.393)‡
Fit-based model enhancements				
Disability ₂ → Disability ₄		0.23 (0.075)**		0.23 (0.076)**
PCS ₁ → PCS ₃		0.33 (0.062)**		--
MCS ₁ → MCS ₄		--		0.13 (0.024)**
LS ₁ → LS ₃		--		0.23 (0.042)**
LS ₁ → Disability ₂		-0.012 (0.003)**		-0.011 (0.003)**
LS ₁ → Disability ₃		-0.014 (0.003)**		-0.016 (0.002)**
LS ₂ → Disability ₃		-0.008 (0.003)*		-0.008 (0.003)*
Indirect effects				
Disability ₁ → Disability ₂ → LS ₃ → PCS/MCS ₄	-0.165 (0.177)	-0.230 (0.111)*	-0.319 (0.121)**	-0.305 (0.096)**
Disability ₁ → LS ₂ → LS ₃ → PCS/MCS ₄	-0.061 (0.066)	-0.166 (0.079)*	-0.120 (0.053)*	-0.178 (0.058)**
Disability ₁ → LS ₂ → PCS/MCS ₃ → PCS/MCS ₄	-0.400 (0.121)**	-0.484 (0.101)**	-0.165 (0.062)**	-0.160 (0.047)**
Total indirect effect	-0.627 (0.242)**	-0.880 (0.187)**	-0.605 (0.170)**	-0.642 (0.137)**
χ^2 / df	640.53/325	477.21/320	628.88/325	461.98/319
RMSEA	0.038	0.027	0.037	0.026
PCLOSE	1.00	1.00	1.00	1.00
CFI	0.982	0.991	0.983	0.992
WRMR	1.127	0.895	1.133	0.893

Notes: All models include constraints where measurement properties (i.e., loadings and thresholds) for the disability measure are equivalent over time (invariant). Measurement paths (i.e., for disability), covariate paths, and covariances between residual terms of downstream endogenous variables within the same wave have been omitted from the table for simplicity, but were included in each model. RMSEA = root mean square error of approximation, PCLOSE = probability that RMSEA ≤ 0.05, CFI = comparative fit index, WRMR = weighted root mean square residual. Standard errors (SE) for indirect effects are from Sobel (1982). ‡*p* = 0.691; †*p* = 0.051; **p* ≤ 0.05; ***p* ≤ 0.01

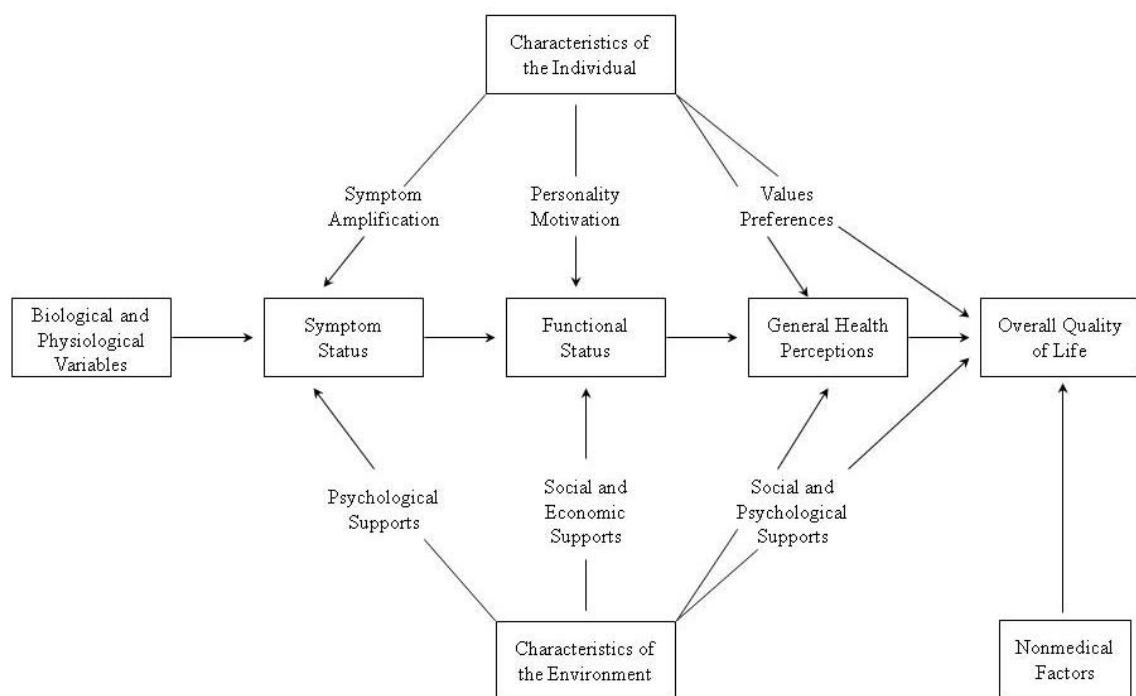


Fig. 1 The Wilson-Cleary health-related quality of life model (Wilson & Cleary, 1995).

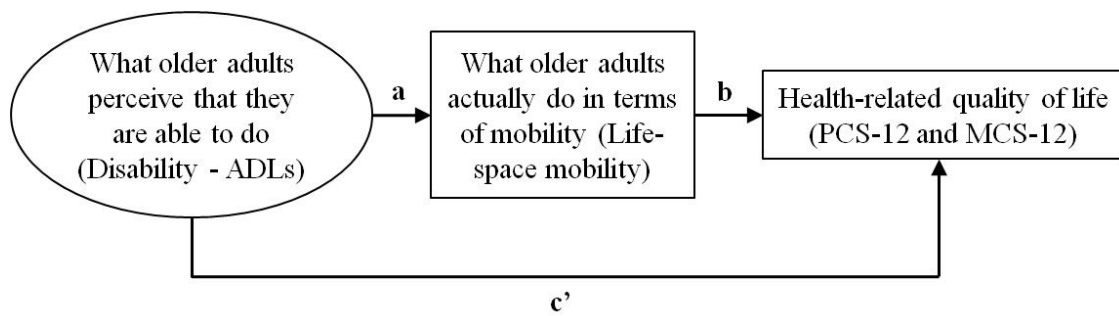


Fig. 2 Basic mediation model for the present study. Indirect effect = ab , direct effect = c' , total effect = $ab+c'$.

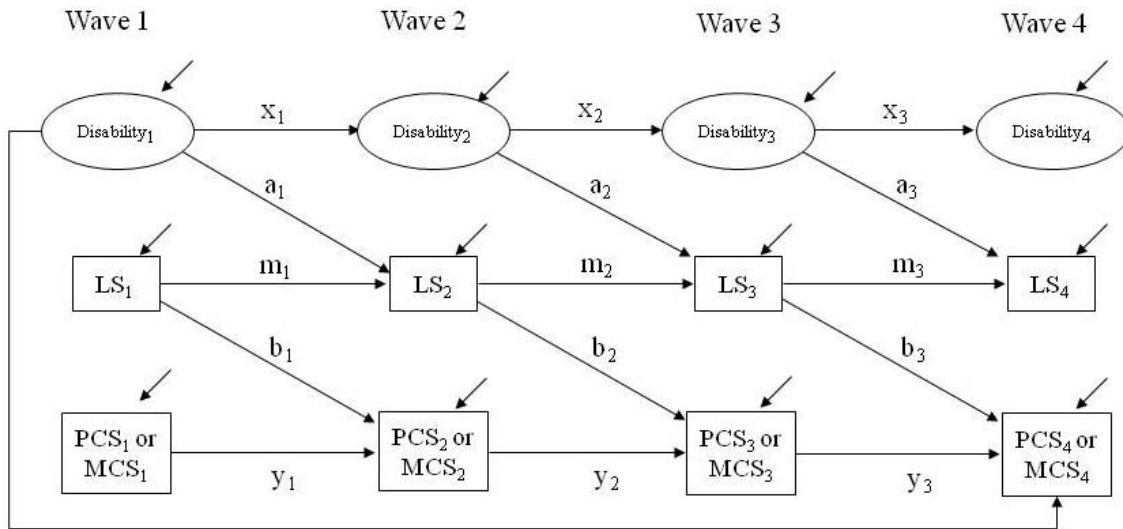


Fig. 3 Autoregressive mediation model framework for the present study. Measurement paths (i.e., for disability), covariate paths, and covariances between residual terms of downstream endogenous variables within the same wave have been omitted for simplicity. LS = life-space mobility and MCS and PCS = mental component summary score and physical component summary score of the SF-12, respectively.

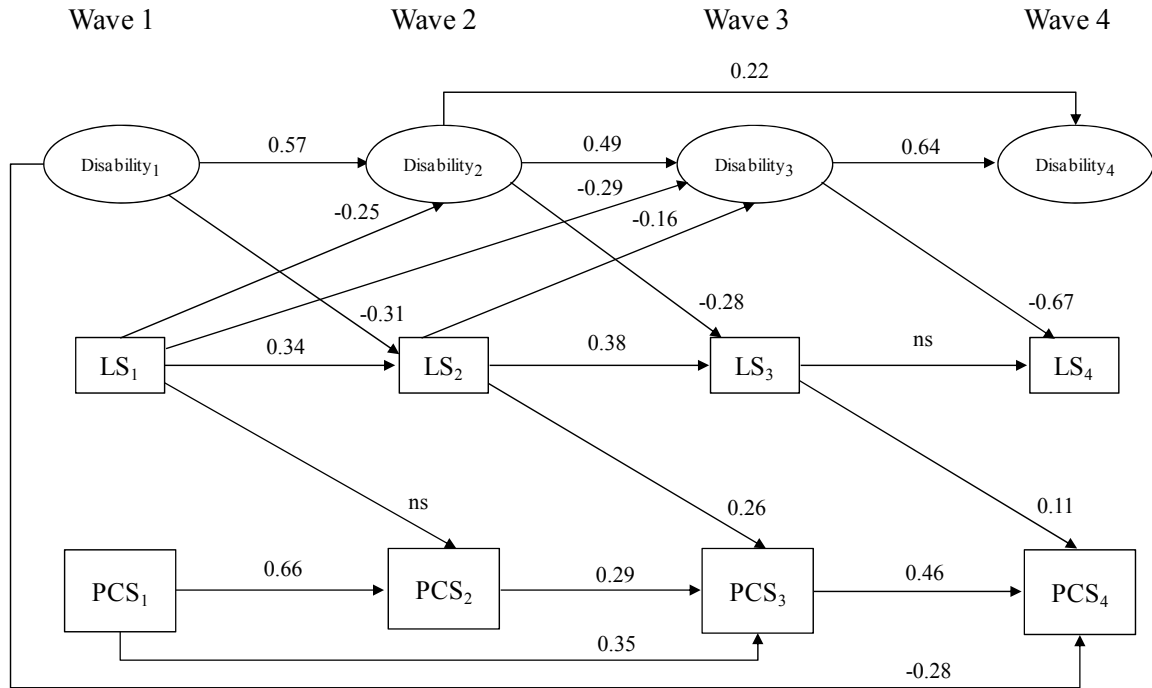


Fig. 4 Completely standardized parameter estimates from the final model (i.e., Model 2) with PCS-12 as the outcome variable including fit-based model enhancements. Measurement paths (i.e., for disability), covariate paths, and covariances between residual terms of downstream endogenous variables within the same wave have been omitted for simplicity. Parameter estimates are located above the corresponding path. LS = life-space mobility and MCS and PCS = mental component summary score and physical component summary score of the SF-12, respectively.

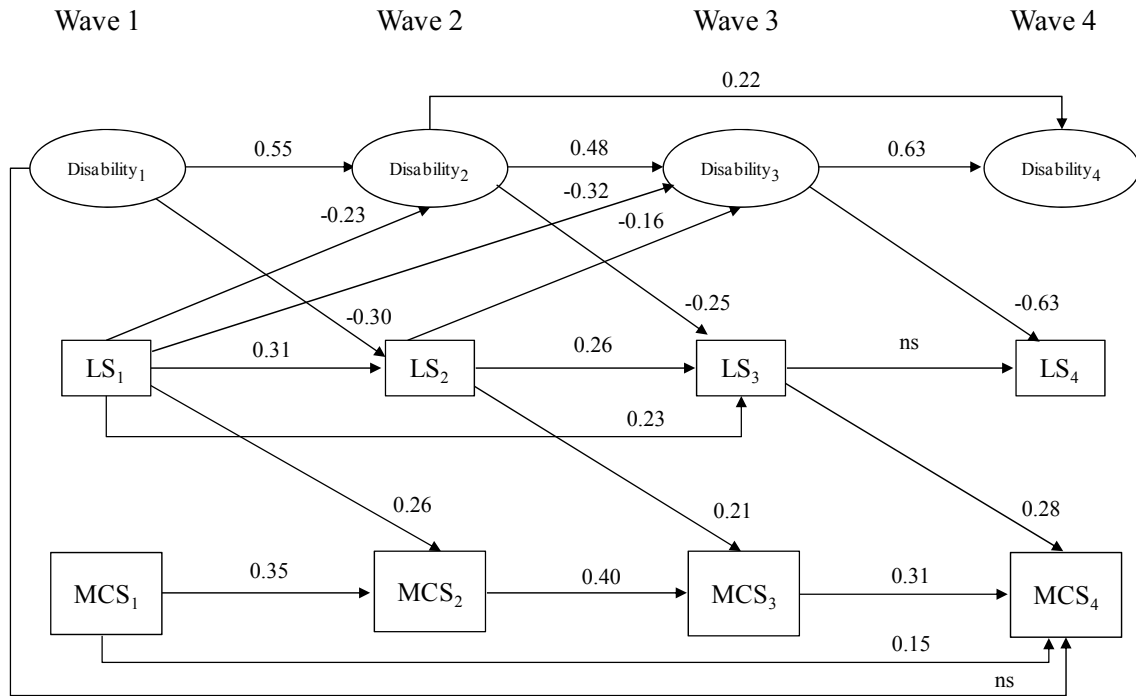


Fig. 5 Completely standardized parameter estimates from the final model (i.e., Model 2) with MCS-12 as the outcome variable including fit-based model enhancements. Measurement paths (i.e., for disability), covariate paths, and covariances between residual terms of downstream endogenous variables within the same wave have been omitted for simplicity. Parameter estimates are located above the corresponding path. LS = life-space mobility and MCS and PCS = mental component summary score and physical component summary score of the SF-12, respectively.

EFFECTS OF RANDOM MEASUREMENT ERROR AND FAILURE TO ACCOUNT
FOR SHARED METHOD VARIANCE IN TESTING AUTOREGRESSIVE
LONGITUDINAL MEDIATION MODELS

by

BENTLEY, J. P., ROTH, D. L., ABAN, I. B., REDDEN, D. T., AND MCCLURE, L. A.

In preparation for *Psychological Methods*

Format adapted for dissertation

Abstract

The autoregressive model is a useful approach for examining mediation, offers several advantages over cross-sectional models of mediation, and is being increasingly used in empirical studies. There are, however, additional complexities associated with this model and no Monte Carlo studies have examined accuracy of parameter estimation and the power for testing mediation in the autoregressive model. A simulation study was conducted to assess the impact on estimation as well as statistical power and Type I error rates of failing to account for random measurement error and shared method variance in the mediator under a variety of conditions, including the degree of shared method variance, degree of composite reliability, degree of stability of the latent mediator, size of the mediated effect, and sample size. The results demonstrate that failure to account for measurement error and shared method variance can have a significant impact on parameter estimation in the autoregressive mediation model, including both overestimation and underestimation of paths of interest. Although the extraction of latent variables from multiple observed measures generally provides accurate estimates and also allows researchers to take into account method effects by allowing correlated measurement errors, latent variable models still require significant levels of composite reliability to achieve acceptable levels of power to detect the mediated effect.

Introduction

At its most basic level, statistical mediation suggests that the causal influence of an independent variable (X) on a dependent variable (Y) is transmitted through another variable called a mediator or intervening variable (M) (Hoyle & Kenny, 1999).

Mediation modeling is important as it attempts to provide mechanistic explanations as to how or why variables exert their influence on other variables. The use of such models to explore the plausibility of theoretical explanations, to evaluate the validity of surrogate endpoints (variables that can be used instead of the ultimate dependent variable), or to assess how or why an intervention produces change in an outcome variable are widely reported in the literature.

A variety of statistical methods for assessing mediation in the context of several different types of study designs have been proposed (e.g., for a review, see MacKinnon, Fairchild, & Fritz, 2007; MacKinnon, 2008a). However, the use of cross-sectional data and a single-mediator model (often with continuous measures of X, M, and Y) is still the norm in most empirical tests of mediation, especially in the field of psychology (Maxwell & Cole, 2007). In the case of true longitudinal mediation, cross-sectional analyses provide biased estimates of the indirect effect in two different models of change, the autoregressive model and a random effects model (Maxwell & Cole, 2007). Assessment of mediation with longitudinal data provides more information regarding the temporal relation of the independent variable, the mediator, and the outcome variable, a

critical component in causal inference of a proposed mechanistic explanation, such as a mediation hypothesis (Kazdin & Nock, 2003). Similarly, causal effects often take time to develop, and variables measured at the same time, as in a cross-sectional analysis, do not account for the necessary development time, essentially assuming immediate effects (Gollob & Reichardt, 1991). In addition, cross-sectional analyses of mediation implicitly assume that prior levels of a variable do not have effects on subsequent levels of that same variable (i.e., autoregressive effects are zero), potentially leading to an omitted variable bias (Gollob & Reichardt, 1991; Cole & Maxwell, 2003). For these reasons, a greater emphasis has been placed on the development of longitudinal mediation models.

There are several classes of models for evaluating longitudinal mediation with the collection of three or more waves of data,¹ including: 1) autoregressive models (Gollob & Reichardt, 1991; Cole & Maxwell, 2003), 2) different types of random effects models, including latent growth curve models (Kenny, Korchmaros, & Bolger, 2003; Cheong, MacKinnon, & Khoo, 2003; Cheong, 2011; von Soest & Hagtvet, 2011), 3) adaptations of latent difference score models which are useful when electing to examine potential differences at different waves of data (Ferrer & McArdle, 2003; McArdle, 2009; Geiser, Eid, Nussbeck, Courvoisier, & Cole, 2010), 4) autoregressive latent trajectory (ALT) models, which combine elements of 1 and 2 (Curran & Bollen, 2001), and 5) a stage-sequential model of mediation based on binary data that is person-centered (i.e., examines response patterns of individuals on the variables of interest), rather than variable-centered (i.e., examines relationships among variables, as in the previous four models) (Collins, Graham, & Flaherty, 1998). Selig and Preacher (2009) provide an

¹ It is possible to provide some evidence of longitudinal mediation with two waves of data. See Cole & Maxwell (2003), MacKinnon (2008b), and Roth & MacKinnon (2012) for examples.

excellent review of the strengths and weakness of the first three models and provide recommendations for appropriate use of each approach.

Concurrent with the growth in methodological developments, several empirical studies have begun to test mediation hypotheses in the context of longitudinal designs with these techniques, primarily autoregressive models (Negri^{ff}, Ji, & Trickett, 2011; Pössel & Thomas, 2011) and latent growth curve models (Cheong, MacKinnon, & Khoo, 2003; Flora, Khoo, & Chassin, 2007; Liu et al., 2009; Audrain-McGovern, Rodriguez, & Kassel, 2009; Roesch et al., 2009; Roesch, Norman, Villodas, Sallis, & Patrick, 2010; Littlefield, Sher, & Wood, 2010). However, there is a need for simulation studies to evaluate the performance of these models under a variety of conditions, evaluating the impact of these conditions on parameter and standard error estimation as well as statistical power and Type I error rates. Cheong (2011) has recently reported on the accuracy of mediated effect estimation and power in the latent growth curve model of mediation. In the present study, a simulation study was conducted to assess the impact of model misspecification (i.e., failing to account for random measurement error and shared method variance in the mediator) in the autoregressive mediation model under a variety of conditions.

Although longitudinal mediation models, such as the autoregressive model, certainly have advantages over the use of cross-sectional mediation modeling, there are limitations and challenges associated with their use. As with the case of cross-sectional models of mediation, the impact of random measurement error in X, M, and Y in an autoregressive model can be substantial (i.e., underestimation of some effects and overestimation of others) and potentially more complex (Cole & Maxwell, 2003; Little,

Preacher, Selig, & Card, 2007). MacKinnon (2008b, p. 209) notes, “One way to improve the interpretability of autoregressive models is to improve measurement of variables either by specifying latent variables or increasing the reliability of measures.”

Specification of latent variables is a commonly used method to address the biasing effect of random measurement error and the autoregressive mediation model can be extended to a latent variable model, where latent variables are extracted from multiple indicators (i.e., observed variables) for X, M, and/or Y.

Despite the availability of latent variable autoregressive models, many investigators continue to use single-item fallible indicators of a construct or linear composites of a set of fallible indicators (e.g., sum scores or average scores of a set of indicators). The impact of such approaches has been evaluated for cross-sectional mediation models (Hoyle & Kenny, 1999; Stephenson & Holbert, 2003; Cheung & Lau, 2008), but not in the autoregressive longitudinal mediation framework. In addition, the use of latent variables in an autoregressive model increases the possibilities of model misspecification, including the failure to account for shared method variance (e.g., the method effects associated with repeated administrations of the same measure), which may also lead to biased parameter estimates (Marsh, 1993; Cole, Ciesla, & Steiger, 2007; Geiser et al., 2010; Kline, 2011). The purpose of this study was to explore the effects of these variables (i.e., the method to handle fallible indicators, omission of paths representing shared method variance) on parameter and standard error estimation as well as statistical power and Type I error rates under a variety of conditions (i.e., degree of shared method variance, degree of composite reliability for the set of indicators representing the mediator, degree of stability of the latent mediator, size of the mediated

effect, and sample size) in the autoregressive mediation model using a Monte Carlo simulation study. Before introducing the study, a brief review of the autoregressive mediation model and a discussion of the random measurement error and shared method variance in the context of longitudinal studies are provided.

The Autoregressive Mediation Model

The basic autoregressive mediation model described by Gollob and Reichardt (1991) and subsequently elaborated on by Cole and Maxwell (2003) can be found in Figure 1 (in this case, there are three measurement occasions or waves). This model involves a single putative causal variable (X), mediator (M), and outcome variable (Y), all three measured at multiple time points. The x , m , and y paths provide evidence of stability of these variables over time (i.e., x , m , and y are regression coefficients summarizing the relationship of a variable with itself at time t and time $t + 1$). The use of single letters without subscripts for paths a and b in Figure 1 suggests that the causal relationship between X and M and M and Y are the same across measurement occasions (i.e., the $X_1 \rightarrow M_2$ path is the same as the $X_2 \rightarrow M_3$ path), a condition known as stationarity.

Cole and Maxwell (2003) outline a set of steps for estimating the autoregressive mediation model with structural equation modeling. The first two steps, testing the measurement model and testing for invariance, require multiple measures for X , M , and/or Y . The third step focuses on the possibility of the omission of important variables (i.e., confounders) by examining correlations among residuals of endogenous variables within each wave (e.g., correlation of the residuals of M_3 and Y_3) (see also Anderson &

Williams, 1992). Significant correlations between such residuals imply that “potentially important variables are missing from the model” (Cole & Maxwell, 2003, p. 571). The fourth step examines the presence of paths that are not part of the proposed mediation model, but may nevertheless bias the estimate of the mediated effect if ignored. These paths include wave-skipping autoregressive paths (e.g., X_3 predicted by X_2 and X_1), cross-lagged relations (e.g., X_2 predicted by M_1 or M_2 predicted by Y_1), and direct effects of X on Y , which, if present, suggests the case of partial, rather than complete, mediation. The fifth step involves the estimation and testing of the mediated (or indirect) and direct effects.

The basic longitudinal mediated effect in Figure 1 (i.e., with the existence of only three waves of data) is $a*b$. Assuming stationarity, which is testable in this model, it should not matter which a or b is used, however it is common to use $X_1 \rightarrow M_2$ as the a path and $M_2 \rightarrow Y_3$ as the b path to reflect temporal ordering. A number of methods exist for the calculation of standard errors and the construction of significance tests and confidence intervals for the mediated effect (see MacKinnon, Lockwood, Hoffman, West, & Sheets (2002) and MacKinnon, Lockwood, & Williams (2004) for a more detailed discussion of the various options available for the standard error of the mediated effect and subsequent construction of confidence intervals). Two commonly used methods for testing the significance of the mediated effect are the joint significance test (Cohen & Cohen, 1983) and the Sobel test (1982). The joint significance test declares a significant mediated effect when the paths comprising the mediated effect are both statistically significant. In the case of the basic autoregressive mediation model in Figure 1, this

would be the $X_1 \rightarrow M_2$ path and $M_2 \rightarrow Y_3$ path. Such a test cannot be used to estimate the mediated effect or in the construction of confidence intervals.

The Sobel standard error formula for the $a*b$ mediated effect is:

$$\sqrt{a^2 s_b^2 + b^2 s_a^2} \quad (1)$$

where s_a and s_b are the standard errors for the a and b estimates. The Sobel standard error is implemented in several statistical software programs and can also be used to construct confidence intervals for the mediated effect, in addition to significance testing. The Sobel standard error is not without its critics. The use of the Sobel standard error relies on asymptotic theory (i.e., normal-theory confidence limits and hypothesis tests). The distribution of a product of two independent normally distributed random variables is generally not normally distributed, although it may approach normality in large samples (MacKinnon, 2008a). The net result is generally conservative hypothesis tests (low Type I error rates and low power) and confidence intervals (i.e., empirical coverage probabilities larger than 95% for a 95% confidence interval).

In a single-mediator, cross-sectional model of mediation, the condition of a zero direct effect (i.e., the relationship between X and Y adjusted for M) implies the case of complete or perfect mediation (Baron & Kenny, 1986). In Figure 1, a direct effect of X_1 on Y_3 is represented by the c' path, such that the condition of $c' = 0$ should imply complete mediation. However, while necessary, this condition is not sufficient for the case of complete mediation in the autoregressive mediation model. Assuming the x and y paths are nonzero, X_1 could directly affect Y_3 (i.e., not go through M) through Y_2 (i.e., $X_1 \rightarrow Y_2 \rightarrow Y_3$ is a time-specific direct effect). In autoregressive mediation models, these

other direct effects must also be zero for mediation to be complete (Cole & Maxwell, 2003).

Random Measurement Error and Shared Method Variance

The impact of random measurement error has received considerable attention in the literature. The impact of fallible measures (i.e., unreliable measures or those possessing random measurement error) is to attenuate true correlations between variables. This biasing effect of measurement error, and its subsequent impact on hypothesis testing, is generally well understood. Because of its role in multiple paths of a mediation model, unreliability in the mediator can lead to a confusing pattern of results (Hoyle & Kenny, 1999; Cole & Maxwell, 2003). In the presence of such unreliability, the standardized $X \rightarrow M$ and $M \rightarrow Y$ paths are underestimated, leading to an underestimation of the indirect effect (the product of these two paths). However, in this situation the $X \rightarrow Y$ path controlling for the mediator (i.e., the direct effect) is actually overestimated. Thus, in some models unreliable variables can attenuate some path coefficients while spuriously inflating others. In the cross-sectional mediation model, these effects have been documented both analytically (Hoyle & Kenny, 1999; Cole & Maxwell, 2003) and through simulation studies in which a linear composite of a set of fallible indicators is used with known levels of composite reliability to represent a mediator, in essence ignoring unreliability (called the observed variable approach) (Hoyle & Kenny, 1999; Stephenson & Holbert, 2003; Cheung & Lau, 2008). Such effects have not been systematically explored in the context of an autoregressive mediation model. The inclusion of prior levels of a variable as a predictor of subsequent levels of the same

variable and the presence of several additional paths may complicate the biasing effects of random measurement error in autoregressive mediation models compared to cross-sectional mediation models (Cole & Maxwell, 2003; Little et al., 2007).

It is possible to extract latent variables from multiple indicators with which to test mediation models. These latent variables are without error and thus their use can substantially minimize the biasing effects of measurement error in mediation modeling (Bedeian, Day, & Kelloway, 1997). The autoregressive mediation model can be extended to a latent variable model (see Figure 2, where a latent variable for M_i is extracted at each of three time points from a set of three indicators).

When assessing the same latent construct at several time points using the same observed measures, it is quite likely that there will be some degree of indicator-specific effect that is present at each testing occasion (Cole & Maxwell, 2003; Raykov & Penev, 2005; Brown, 2006). This is a type of shared method variance and is often reflected by correlations between residuals (i.e., measurement errors or disturbances) of the same indicator across multiple time points (i.e., cross-wave, within-construct error covariance). This is noted in Figure 2 by the doubled-headed arrows between the error terms of indicators at multiple time points (e.g., $m_{A1} \leftrightarrow m_{A2}$). This presence of correlated measurement errors suggests another exogenous common cause, in this case a common method, in addition to the latent factor, M . These effects are usually not of direct interest, but failure to account for them may lead to poor overall model fit and to biased parameter estimates (Marsh, 1993; Cole, Ciesla, & Steiger, 2007; Geiser et al., 2010; Kline, 2011). In the case of the autoregressive mediation model, ignoring correlated measurement errors can result in an overestimation of construct stability (i.e., the x , m , and y paths in

Figures 1 and 2). Approaches such as creating linear composites for a construct such as a mediator from a set of fallible indicators ignore both measurement error and the possible presence of shared method variance. The consequences for such an approach are complex in the autoregressive mediation model since disregarding measurement error has a downward biasing effect on stability coefficients while failure to specify correlated measurement errors has an upward biasing effect.

The concept of reliability discussed when several items are used to measure a construct is referred to as internal consistency reliability, most frequently summarized by Cronbach's alpha (although there are other measures available). This is not equivalent to a reliability stability coefficient, which is often assessed by the correlation (e.g., Pearson's or intraclass correlation coefficients) between sets of scores from the same subjects on the same measure administered at two different times. The x , m , and y paths in Figures 1 and 2 can be thought of as stability coefficients, albeit the m and y paths reflect stability in M or Y after adjusting for X or M , respectively. In some areas, notably the behavioral sciences, highly stable constructs are sometimes labeled as "traitlike," whereas unstable variables (those with low stability coefficients) are labeled as "statelike" (Cole, Martin, & Steiger, 2005). Given the focus of the present study is on unreliability and correlated measurement errors with respect to the mediator in an autoregressive mediation model, the degree of stability for this construct needs to be taken into consideration.

Methods

Population Model

The population longitudinal mediation model was based on the model in Figure 2, a three-wave model with measures for X, M, and Y at all three measurement occasions. Multivariate normal data with known population values were generated as part of the simulation. The X and Y variables were assumed to be observed variables and a total of three observed indicators were generated for the latent variables M_1 , M_2 , and M_3 . The X and Y variables as well as the indicators of M were assumed to be continuous and the X and Y variables were assumed to be measured without error. Based on the work of Cheung (2007), for ease of manipulation, population values were selected such that the X and Y observed variables and the latent variables representing M at all three waves were standardized (mean = 0; standard deviation = 1).

The mediated effect was defined as the independent variable at time 1 (X_1) affecting the dependent variable at time 3 (Y_3) via the mediator at time 2 (M_2) and was estimated as the product of these two paths in all subsequent analyses ($a*b$). No contemporaneous relationships (i.e., $X_2 \rightarrow M_2 \rightarrow Y_2$) nor cross-lagged relationships (i.e., $M_1 \rightarrow X_2$) were specified (all assumed to be zero). Population values for all direct effects of X_1 on Y_3 were assumed to be zero (i.e., complete mediation). The covariances among residual terms of endogenous variables within each wave were also constrained to be zero. This suggests no important variables are missing from the model.

Population stability coefficients for X and Y (i.e., the x and y paths) were the same for all conditions ($x = 0.71$ and $y = 0.51$). The paths selected represent about 50% of the variance in X_2 and X_3 under all experimental conditions and about 26% of the variance in Y_2 and Y_3 under the conditions of no mediated effect. The a , b , and m path

coefficients were manipulated factors in the simulation study. Within a given experimental condition, all paths (a , b , x , m , and y) were set to be invariant over time in the population model (e.g., the $M_1 \rightarrow M_2$ path = $M_2 \rightarrow M_3$ path and the $X_1 \rightarrow M_2$ path = the $X_2 \rightarrow M_3$ path). Furthermore, the condition of equilibrium in the population model was assumed, meaning that the correlations among X , M , and Y are the same at all three measurement occasions. This feature was used to determine the correlations among X , M , and Y at time 1 under the different experimental conditions (see Appendix A).

Simulation Study Conditions

The simulations and subsequent model fitting were conducted using *Mplus* 6.11 (Muthén & Muthén, Los Angeles, CA). The data generation source code for one of the combinations of the experimental conditions along with code for the fitting of the three model specifications can be found in Appendix B. A total of six factors were manipulated in this simulation study: 1) three model specifications, 2) three levels of composite reliability for the set of indicators representing the mediator, 3) three levels of shared method variance, 4) two levels of stability of the latent mediator, 5) four sizes of the mediated effect, and 6) four different sample sizes. Of these, only five contributed to the creation of different sets of replications (model specification can be thought of as a within-subjects factor, where all three levels of this factor were applied to all replications in each condition). Thus, data were generated under 288 unique experimental conditions ($3 \times 3 \times 2 \times 4 \times 4$).

A total of 600 replications were simulated for each of these 288 conditions and a model outlined in Figure 2 (including the six measurement error covariances, the c_1 ' and

c_2' paths, and the loading of the first indicator for the mediator at each of the three waves fixed to 1 to set the scale for the constructs and identify the model) was fit to the 600 raw datasets in each condition using maximum likelihood estimation. Out of these 600 replications in each of the 288 conditions, replications that failed to converge or resulted in negative variances or residual variances were excluded, and then the first 500 replications with converged and proper solutions were selected for subsequent analyses (Paxton, Curran, Bollen, Kirby, & Chen, 2001), providing a total of 144,000 data sets. No experimental condition had more than 100 replications with improper solutions and in most conditions, all solutions were proper. Almost all of the problematic replications occurred with $n = 100$ and low reliability of the mediator. The three specified models described below were then fit to these 144,000 simulated data sets.

Model specification. Three different models were specified and were fit to each data set. The first model was based on Figure 2 and included the extraction of latent variables for M_1 , M_2 , and M_3 in addition to the specification of the six measurement error covariances noted in the figure (i.e., these covariances were freely estimated). Under the conditions where there were nonzero covariances between the error terms of the indicators at multiple time points, this was the properly specified model. When these covariances were zero (i.e., no shared method variance), this model included an error of inclusion. The second model specification also included the extraction of latent variables for the mediator at all three measurement occasions, but failed to specify indicator error covariances (i.e., these covariances were not estimated, but rather fixed to zero). Thus, under conditions of no shared method variance, this was the properly specified model, whereas it served to illustrate the potential biasing effects of ignoring correlated

measurement errors when these errors were present in the population model. The third model specification essentially failed to explicitly acknowledge unreliability in the mediator or the possible presence of shared method variance. In this model, an unweighted linear composite was created for the mediator variable at each measurement occasion by averaging the values of the three indicators at each time point.

For all three models, the c_1' and c_2' paths (i.e., direct effects) were estimated to assess the potential for overestimation of direct effects under various conditions. In addition, although all paths (a , b , x , m , and y) were set to be invariant over time in the population model, no such constraints were imposed during model fitting (i.e., all autoregressive model parameters were freely estimated). Maximum likelihood was used as the estimation method.

Composite reliability for the mediator. Three levels of composite reliability of the mediator were examined: 0.60, 0.75, and 0.90, representing low, moderate, and high reliability. Factor loadings for each indicator were set to 1 such that reliability was manipulated by changing the residual (or error) variances of the factor indicators. The composite reliability coefficient can be expressed as (Fornell & Larcker, 1981):

$$\rho = \frac{(\sum \lambda_i)^2}{(\sum \lambda_i)^2 + \sum \theta_i} \quad (2)$$

where $(\sum \lambda_i)^2$ is the squared sum of unstandardized factor loadings and $\sum \theta_i$ is the sum of unstandardized residual variances for the indicators. Given factor loadings of 1, residual variances of 2, 1, and 1/3 for the indicators provided composite reliabilities of 0.60, 0.75, and 0.90, respectively. Thus, simplifying assumptions that the loadings and residual variances for the indicators are the same within wave (such an approach was also

used by Hoyle and Kenny (1999) and are also equivalent across waves were used (i.e., time invariance of the measurement model). Although this was the case in the population used to generate sample data, no such constraints were imposed during model fitting.

Shared method variance. Three levels of shared method variance, reflected by covariances between the error terms of the same indicators at multiple time points (e.g., $m_{A1} \leftrightarrow m_{A2}$), were tested. Correlations of 0, 0.10, and 0.30 among measurement errors of a given indicator at adjacent time points were chosen to reflect no shared method variance, a small amount, and a medium amount using Cohen's (1988) guidelines for the correlation coefficient. As the covariance between measurement errors is a function of the correlation coefficient and the residual variances of the indicators, the covariance was larger in conditions in which the reliability was lower (i.e., the residual variance for the indicators was larger). Only covariances between adjacent time points were specified (see Figure 2) and the same value was used for all six possible correlations. While there were correlated errors across time for the same indicator, no correlated measurement errors were specified within the construct on the same measurement occasion. This indicates no occasion-specific effects and reflects common practice.

Stability of the latent mediator. Two levels of stability in M , after adjusting for other relations in the model (i.e., path m in Figure 2), were manipulated: 0.36 and 0.51. These represent about 13% and 26% of the variance in M_2 and M_3 under the conditions of no mediated effect.

Size of the mediated effect. Four levels of a and b were used to vary the effect size of the mediated effect. For ease of manipulation, population path coefficients for a and b were set to be equal (see Cheung, 2007; 2009, for similar designs). The values of 0, 0.10,

0.25, and 0.34 were chosen for the a and b paths to reflect zero, small, medium, and large effects for the mediated effect. These values correspond to zero, small (0.02), medium (0.15), and large (0.35) effects on the f^2 metric for the incremental effect in the variance explained in Y_3 of including the a and b paths in the autoregressive model over the y path when $m = 0.36$ (Cohen, 1988) (see Appendix A for further elaboration). When $m = 0.51$, the effects are slightly larger, but still within Cohen's guidelines for small, medium, and large effects.

Sample size. Four sample sizes often encountered in social and behavioral science research were considered: 100, 200, 500, and 1000.

Study Outcomes

Parameter and standard error estimation. Both bias and relative bias were used to assess the accuracy of point estimation in the autoregressive mediation model. While the primary interest was the estimate of the mediated effect (i.e., the product of the $X_1 \rightarrow M_2$ and the $M_2 \rightarrow Y_3$ paths or $a*b$) and the direct effects (i.e., the $X_1 \rightarrow Y_2$ and the $X_1 \rightarrow Y_3$ paths), it is important to look at the effect of model misspecification on the estimation of other parameters in the model as it may help to explain the bias in the estimates of the mediated and direct effects. Relative bias is defined as the ratio of the deviation in an estimate from the true value to the true value:

$$\frac{\hat{\beta} - \beta}{\beta} \quad (3)$$

$\hat{\beta}$ is the point estimate of the effect of interest from the simulated data whereas β refers to the true value specified in the population model. Absolute values greater than 0.10

were considered problematic (Kaplan, 1988). Under conditions where the true value was zero, the bias (the numerator in equation 3) was calculated and reported.

Because of questions concerning its performance especially with smaller samples, the relative bias associated with the estimation of the Sobel (1982) standard error is often reported in simulation studies of mediation (e.g., see MacKinnon et al., 2002; Taylor, MacKinnon, & Tein, 2008; Cheong, 2011). To calculate this relative bias, equation 1 was used to calculate the Sobel (1982) standard error estimate for each of the 500 replications under a given set of conditions. The standard deviation of the mediated effect across the same 500 replications was used as the true value of the standard error and then equation 3 was used to calculate relative bias.

Power and Type I error rates. Both the mediated and the direct effects were tested for significance in each of the 500 replications under all experimental conditions using a two-tailed test at the $\alpha = 0.05$ level of significance. Both the Sobel test and the joint significance test were used for testing the mediated effect. Under conditions when the true mediated effect was 0 and for all tests of the direct effects (because the true values of the c_1' and c_2' paths in Figure 2 are 0 under all conditions), the proportion of replications in which the null hypothesis was rejected was used as an estimate of the Type I error rate. When the true mediated effect was nonzero, the proportion of replications where the mediated effect was statistically significant was used as the measure of power.

Analysis

SAS 9.2 (SAS Institute, Cary, NC, USA) was used for analyzing the simulation results. To examine the main effects and interactions of the six design factors on the study outcomes, analysis of variance was used for continuous outcomes (bias and relative bias) and logistic regression was used for dichotomous outcomes (Type I error and power). Because the number of observations was so large, effects were interpreted on the basis of effect sizes rather than conventional $p < 0.05$ significance testing (Paxton et al., 2001). The proportion of total variation accounted for (η^2) was used as the measure of effect for ANOVA models and the proportional reduction in deviance attributable to a predictor (R_L^2) was used for logistic regression models. The number of observations coupled with the goal of estimating terms for all possible interactions precluded the use of categorical predictors for all factors in the logistic regression models (i.e., the model could not be estimated). Thus, these models were estimated with all factors except model specification treated as centered, quantitative predictors (see Taylor et al., 2008 for a similar approach). Although the three model specifications comprise a within-subjects factor, model specification was treated as a between-subjects factor in the analysis. Because the results were interpreted in terms of effect sizes rather than significance testing, the treatment of model specification as a between-subjects factor did not affect the results.

Results

When the three specified models were fit to the 144,000 simulated data sets, proper solutions were obtained for all latent variable models with indicator error

covariances specified (as this model was used to initially assess for nonconvergence and out-of-range values, this was expected) and for all composite variable models (i.e., models in which a linear composite was created for the mediator variable at each measurement occasion by averaging the values of the three indicators). However, estimation problems were encountered in 205 data sets (less than 0.1% of the total number of data sets) when the third model specification was used (i.e., latent variable models with indicator error covariances unspecified) and these were excluded from subsequent analyses, leading to a slightly unbalanced design. The most frequently encountered problem for this model specification was a negative residual variance for one of the indicators and almost all occurred with $n = 100$ and low reliability of the mediator.

Parameter Estimation

Overall, the study factors (and all possible interactions) accounted for 3.3% of the total variation in the relative bias associated with estimates of the mediated effect. Most of the explained variation (93%) was due to the main effects of, and interactions among, three factors: model specification, level of reliability, and degree of indicator error correlation (i.e., shared method variance). Mean relative bias values for the mediated effect estimates are presented in Table 1 as a function of these three factors, as well as sample size, collapsing across levels of latent mediator stability and size of the mediated effect. Even though sample size or interactions with this factor did not account for nontrivial variance, results are nevertheless reported separately for this factor as it helps to demonstrate that the biasing effects of model misspecification do not significantly vary by sample size.

Although there was a slight degree of positive bias at small sample sizes paired with low to moderate reliabilities for the latent variable model when indicator covariances were estimated, the relative bias for the mediated effect was less than 0.10 in all conditions with this model specification. When the indicator covariances were unspecified in a latent variable model, the mediated effect was significantly underestimated in conditions of low to moderate reliability coupled with a medium amount of shared method variance, a finding that did not change with increasing sample sizes. Finally, in conditions in which the composite variable model was fit, the absolute value of the relative bias of the mediated effect was almost always larger than 0.10. Higher reliability was generally associated with a smaller amount of bias, and within each level of reliability, greater amounts of indicator error correlation appeared to increase the degree of bias. Again, increasing sample size did not attenuate the biasing effects of this model misspecification.

Because relative bias could not be calculated when the mediated effect was zero, a separate analysis examined the bias of the mediated effect estimates under this condition. None of the study factors nor interactions accounted for a nontrivial amount of variation (0.2%) and the bias values were all generally near zero, suggesting that ignoring shared method variance and failing to explicitly recognize unreliability in the mediator do not appear to lead to biased estimates of the mediated effect when the effect was zero (not tabled).

The patterns of means for the relative biases associated with estimates of the m , a , and b paths can be found in Tables 2, 3, and 4, respectively. For all three of these outcomes, the study factors accounted for a significant amount of total variation in

relative bias (29.8%, 2.5%, and 10.5%, respectively) with most of the explained variation (97%, 85%, and 97%, respectively) coming from the main effects of, and interactions among, model specification, level of reliability, and degree of indicator error correlation. Thus, the results in Tables 2, 3, and 4 are presented as a function of these three factors, as well as sample size, similar to Table 1. Because the mediated effect was defined as the product of the $X_1 \rightarrow M_2$ (a path) and the $M_2 \rightarrow Y_3$ (b path) paths, only results for these two paths are presented in Tables 3 and 4. However, the results for the other a ($X_2 \rightarrow M_3$) and b ($M_1 \rightarrow Y_2$) paths were generally the same. Likewise, Table 2 only presents results for the first m path ($M_1 \rightarrow M_2$), as results for the other m path followed basically the same pattern.

For path m , an interesting pattern emerges when comparing the different model specifications (Table 2). Consistent with the findings for the mediated effect, the latent variable model with specified indicator covariances generally performed well, with a small amount of positive bias at small samples sizes with low to moderate amount of reliability. In the latent variable model with unspecified error covariances, the m path is increasingly overestimated as the degree of shared method variance increases, an effect that diminishes as reliability increases. This same path is generally underestimated in the composite variable model, but since this model specification does not explicitly acknowledge unreliability or the presence of shared method variance, the relative bias tends toward zero as the degree of shared method variance increases, reflecting the upward biasing effect of ignoring shared method variance.

Results for path a estimates are generally inversely related to those of path m (Table 3). Namely, as path m is underestimated by a given model, path a is generally

overestimated, and vice-versa. These results are best seen in the two models with notable misspecifications, as the latent variable model with indicator error covariances specified generally has values for relative bias for path a near zero. As with other paths, the biasing effects of model misspecification tend to diminish as reliability increases.

As seen with the other paths and the mediated effect, path b tended to be slightly overestimated in the latent variable model with specified indicator covariances at small samples sizes with low to moderate amount of reliability (Table 4). When the indicator covariances were unspecified in a latent variable model, path b was underestimated in conditions of low and a medium amount of shared method variance. Finally, for the composite variable model, a fairly consistent pattern emerged: a negative bias for this path that improved with increasing reliability, but was invariant to both sample size and shared method variance within each level of reliability.

Paths a and b were zero when the mediated effect was zero, thus relative bias could not be calculated under these conditions. Not surprisingly, separate analyses examining the bias associated with estimates of these paths under these conditions revealed similar results as the bias of mediated effect estimates. Namely, none of the study factors nor interactions accounted for nontrivial amount of variation (0.1% for the a path and 0.2% for the b path) and the bias values were all generally near zero (not tabled).

With respect to estimates of the two direct effects, $X_1 \rightarrow Y_2$ and $X_1 \rightarrow Y_3$, the study factors taken together accounted for 3.5% and 7.1% of the total variation in the bias, respectively (relative bias could not be calculated as the true values of these paths were zero). The main effects of, and interactions among, model specification, level of

reliability, and effect size of the mediated effect accounted for most of the explained variation in the bias associated with these direct effects (93% and 94%, respectively). Tables 5 and 6 present mean bias values for the direct effects as a function of these three factors, as well as sample size, collapsing across levels of latent mediator stability and degree of indicator error correlation. Table 5 shows that the positive bias in the $X_1 \rightarrow Y_2$ path primarily occurred in the composite variable model with lower levels of reliability coupled with larger mediated effects; the latent variables models generally showed no significant bias associated with this path. Similarly, the $X_1 \rightarrow Y_3$ path tended to be overestimated in the composite variable model with low to moderate levels of reliability when the mediated effect size was medium to large. There was a small degree of positive bias for this path in the latent variable model with unspecified indicator error covariances at low levels of reliability and large mediated effects. Although the degree of shared method variance or interactions with this factor did not account for a nontrivial amount of variation in this outcome, it is worth noting that this positive bias generally occurred in conditions where the degree of shared method variance was higher.

The study factors generally accounted for only a small amount of variation in the relative bias associated with estimates of the y paths (<2%). Furthermore, an examination of cell means revealed that although there tended to a slight positive bias for the composite variable model at low to moderate levels of reliability couple with large mediated effects, none of the conditions produced relative biases greater than 0.10.

Standard Error Estimation

The relative biases of the Sobel standard error of the mediated effect can be found in Table 7. Because the main effects of, and interactions among, model specification, level of reliability, effect size of the mediated effect, and sample size accounted for most of the explained variation (86%, with total variation accounted by all study factors of 7.0%), the results in Table 7 are collapsed across levels of latent mediator stability and degree of indicator error correlation. The size of the mediated effect had by far the largest effect, which is evident in Table 7, which shows that the Sobel standard error tended to be positively biased under all model specifications and at all sample sizes, when the size of the mediated effect was zero (recall that both paths comprising the mediated effect were zero under this condition). When the mediated effect was nonzero, the absolute value of the relative bias of the Sobel standard error estimator was generally below 0.10 and improved with increasing sample size.

Type I Error and Power

The Type I error rates for the mediated effects (i.e., when the true mediated effect was zero) using both the Sobel test and the joint significance are reported in Table 8. Because rejecting a true null hypothesis was such a rare event under all conditions, no assessment was made of the contribution of the study factors. Results in Table 8 are presented by model specification and sample size as well as the method of testing (i.e., Sobel and joint significance tests). Type I error rates were well under the nominal level of 0.05 for all sample sizes, regardless of model specification or method of testing (although Type I errors rates were generally slightly higher for the joint significance test).

The results with respect to empirical power for the mediated effect can be found in Table 9. The study factors accounted for a 62.4% deviance reduction for the Sobel test and 55.9% for the joint significance test, with most of the contribution to this reduction (99.9% for both tests) coming from the main effects of, and interactions among, model specification, level of reliability, effect size of the mediated effect, and sample size. Table 9 therefore presents empirical power levels for different levels of these factors. In general, regardless of model specification and method used to test mediation, power generally increased with increasing sample size, increasing effect size of the mediated effect, and increasing reliability of the mediator, as would be expected. There was also a power advantage associated with the joint significance test relative to the Sobel test. Although specifying indicator error covariances did not appear to significantly affect the power of the test for the mediated effect, there was a modest reduction in power in general associated with the two latent variable modeling approaches relative to the composite variable model.

Finally, Table 10 provides the Type I error rates associated with the test of the two direct effects, $X_1 \rightarrow Y_2$ and $X_1 \rightarrow Y_3$. For each outcome, the study factors accounted for a nontrivial deviance reduction, 4.2% and 10.6%, respectively. Most of these effects (93% and 95%, respectively) were attributed to model specification, level of reliability, effect size of the mediated effect, and sample size (and their interactions). Thus, the entries in Table 10 are mean values collapsed across the other two factors (i.e., degree of indicator error correlation and latent mediator stability). Type I error rates for the direct effects generally followed the bias in the direct effects presented in Tables 5 and 6. Thus, the Type I error rates significantly exceeded nominal levels (0.05) generally in the

composite variable model, especially with lower levels of reliability coupled with larger mediated effects. As expected, Type I errors for the direct effects increased in frequency with increases in sample size. Under conditions of low reliability, a large mediated effect, and a sample size of $n = 1,000$, the Type I error rate for the $X_1 \rightarrow Y_3$ direct effect was 0.811 in the composite model.

Discussion

The autoregressive model is a useful approach for examining mediation and offers several advantages over cross-sectional models of mediation, most notably is that it explicitly recognizes the temporal relation among variables involved in a hypothesized causal relationship and allows for the control of the potential confounding effects of prior levels of a variable when assessing mediation relationships. However, there are additional complexities associated with the autoregressive model of mediation. The effects of measurement error, while well documented in cross-sectional models of mediation (Hoyle & Kenny, 1999; Stephenson & Holbert, 2003; Cole & Maxwell, 2003; Cheung & Lau, 2008), are potentially more complicated in the autoregressive mediation model because of the addition of multiple paths and the control of prior levels of the X, M, and Y variables. Cole & Maxwell (2003, p. 568) note that “under such circumstances, the biasing effects of measurement error become utterly baffling.” In addition, the use of the same observed measures at multiple time points in the autoregressive model increases the possibility of indicator-specific variance, a type of shared method variance that can create estimation problems if ignored. Beyond these two additional complexities, to our knowledge, no Monte Carlo studies have been published on the power for testing

mediation in the autoregressive model. Therefore, this study examined the effects of failing to account for random measurement error and shared method variance in the autoregressive mediation model and also provided information concerning the power of the autoregressive mediation model under a variety of conditions.

In general, a latent variable approach, where covariances were specified between measurement errors of a given indicator at adjacent time points, performed well in terms of estimating the parameters in an autoregressive mediation model, including the mediated effect, under all study conditions. However, in situations when the mediator is not measured reliably, caution should be exercised when sample sizes are less than 200. Failing to estimate nonzero error covariances led to underestimation of the mediated effect as the degree of error correlation (or shared method variance) increased, especially in conditions of low reliability. Greater sample sizes did not alleviate this biasing effect due to model misspecification. The primary mechanism for this effect is an overestimation of the stability coefficient for the mediator with consequent underestimation of the two paths comprising the mediated effect (i.e., paths a and b). The consequences of failing to account shared method variance had minimal impact on the estimation of the direct effects.

A composite variable model, which combines multiple measures into a single, unweighted score, essentially ignores both unreliability as well as shared method variance when it is present. Not surprisingly, this approach had a complicated pattern of biasing effects on the autoregressive mediation model parameters under the various study conditions. The composite variable approach generally led to an underestimation of the mediated effect, consistent with cross-sectional simulation studies (Hoyle & Kenny,

1999; Stephenson & Holbert, 2003; Cheung & Lau, 2008), especially when the reliability of the mediator was low in the presence of moderate levels of shared method variance. While measurement error in the mediator attenuates the stability coefficient, m , ignoring shared method variance creates a biasing effect in the other direction. Thus, it appears that the relative bias in m *improves* in the presence of increasing shared method variance in the composite variable model, which is misleading at best. Furthermore, as the relative biases for path a and path m generally move in opposite directions, the positive bias for path a generally decreased in the composite variable model as the degree of shared method variance increased. Given that the relative bias of path b was relatively constant in a negative direction at a given level of reliability for the composite variable model, this explains the pattern of bias for the mediated effect. As seen in cross-sectional models of mediation (Hoyle & Kenny, 1993; Cole & Maxwell, 2003), when the mediator is measured with error, the direct effects are spuriously inflated (with consequent effects on the probability of incorrectly rejecting the null hypothesis of no direct effect), especially in the presence of larger mediated effects. In some cases, ignoring unreliability and the presence of shared method variance appear to “cancel out” each others’ biasing effects, while in other cases, namely the mediated effect, the biasing effects are compounded. This highlights the multifaceted effects of unreliable variables in the autoregressive mediation model; these effects that are multiplied when one considers potential measurement errors in X and Y in addition to M .

The present model only examined three waves of data. The addition of four or more waves of data further complicates the biasing effects of measurement error and the failure to account for shared method variance. With three waves of data, the mediated

effect of interest is simply $a*b$, but with more than three waves, one can estimate both time-specific indirect effects, as well as the overall indirect effect (Gollob & Reichardt, 1991; Cole & Maxwell, 2003; Selig & Preacher, 2009). For example, for a four-wave, autoregressive mediation model, there are three time-specific indirect effects (assuming stationarity):

1. $X_1 \rightarrow X_2 \rightarrow M_3 \rightarrow Y_4 = x*a*b$
2. $X_1 \rightarrow M_2 \rightarrow M_3 \rightarrow Y_4 = a*m*b$
3. $X_1 \rightarrow M_2 \rightarrow Y_3 \rightarrow Y_4 = a*b*y$

The overall indirect effect of $X_1 \rightarrow Y_4$, or the change in Y_4 due to X_1 that is mediated by M , is the sum of the time-specific indirect effects. There are six time-specific indirect effects in a five-wave model. These mediated effects involve the product of several paths, all of which themselves can be attenuated or spuriously inflated in the presence of measurement error or when failing to account for true shared method variance. Thus, there are many more ways in which one can be wrong in these situations and a true estimate of the mediated effect becomes almost impossible in a multiple-waves autoregressive mediation model when unreliable variables are used and shared method variance is not addressed.

Although the latent variable approaches with specified indicator error covariances provided significantly better parameter estimates in terms of bias, they did lead to a modest reduction in statistical power for the mediated effect compared to the composite variable model, primarily because the standard errors tend to be higher in latent variable models because of the estimation of more parameters. When using unreliable variables, composite variable models tend to provide biased estimates (of not just the mediated effect, but direct effects as well) with smaller standard errors as opposed to unbiased

estimates with larger standard errors in the case of latent variable models. Similar findings have been found with cross-sectional models of mediation (Hoyle & Kenny, 1999). Although the biasing effects of measurement error can generally be corrected by extracting latent variables from multiple observed variables, it is important to recognize that simply obtaining multiple measures and extracting a latent variable is not enough. Even in latent variable models, power can be significantly enhanced by using a set of measures with a high degree of composite reliability. Sample sizes of 200 or larger are generally sufficient for 0.80 power in the autoregressive mediation model, given a mediator of moderate to high reliability and a medium to large mediated effect. For conditions of low or moderate reliability in the mediator, even sample sizes of 1,000 are generally not large enough to achieve 0.80 power to detect a small mediated effect.

As has been shown in a variety of other simulation studies of mediation for different study designs, the joint significance test tends to be more powerful than the Sobel test (MacKinnon et al., 2002; Taylor et al., 2008; Cheong, 2011). Both methods had Type I error rates well below 0.05. Other methods of testing, such as other methods of calculating standard errors of the mediated effect (e.g., see MacKinnon et al., 2002), one of several methods of bootstrapping (e.g., see Shrout & Bolger, 2002; MacKinnon et al., 2004; Cheung & Lau, 2008; Taylor et al., 2008), or methods that construct asymmetric confidence intervals based on the theoretical distribution of the product of two random normal variables (MacKinnon et al., 2004, MacKinnon, Fritz, Williams, & Lockwood, 2007), have been assessed in simulation studies of the mediated effect in a number of different study designs. Future studies using these methods in an autoregressive mediation modeling framework are warranted.

Investigators should be encouraged to consider the possibility of the presence of correlated measurement errors among the same indicators that are repeatedly measured. Some investigators have included measurement error correlations without a priori specification as a method to improve model fit, a practice that has led some to caution against the use of correlated errors in structural equation modeling. However, this should not cause researchers to overlook the inclusion of theoretically justified or design-driven correlated residuals (Cole et al., 2007; Kline, 2011) such as those commonly present in longitudinal research. In the present study, specifying indicator error covariances in all situations, even when such correlations were not present in the true population, did not have any biasing effects on the parameters of interest, nor did it appear to significantly affect the power of the test for the mediated effect. Based on this, one could conclude that researchers should always specify such correlated residuals in longitudinal studies. However, it is possible that such specifications may lead to models that fail to converge or cases of out-of-range values such as negative variances/residual variances. Furthermore, one may have theoretical justification for not including such correlated errors, such as the case of designs that employ wide assessment intervals, where such effects may be less likely (Brown, 2006).

It is also important to recognize that there are other approaches to account for shared method variance other than correlating the residuals of indicators. Geiser et al. (2010) label this approach the “multi-occasion correlated uniqueness” (CU) approach and note that this approach has some limitations, most notably that the number of estimated correlations can increase substantially with many indicators, constructs, and time points. The approach advocated by Geiser et al. (2010) is one in which indicator-specific

variance is captured by an indicator-specific (IS) factor rather than a set of correlated errors (see also Raffalovich & Bohrnstedt, 1987; Eid, Schneider, & Schwenkmezger, 1999). Future research is needed to evaluate whether the use of IS factors can significantly enhance estimation accuracy and statistical power in the autoregressive mediation model.

Although the current findings provide researchers with practical information to consider when assessing mediation in the context of an autoregressive model, this study is not without its limitations. First, although this study did assess Type I error rates for tests of the mediated effects, the only zero-mediated effect condition tested the case when both paths comprising the mediated effect were zero. Several studies have shown that both the relative bias of the Sobel standard error and the Type I error rates associated with significance tests of the mediated effects can vary when one or more paths comprising the zero mediated effect has a true nonzero value (MacKinnon et al., 2002; Taylor et al., 2008). Second, an assumption of the present analysis is that the waves are separated by the optimal time interval. The timing and spacing of measurements are critical issues for researchers designing longitudinal studies as estimation and the power to detect effects can be significantly influenced by these design considerations (Gollob & Reichardt, 1991; Collins and Graham, 2002; Cole & Maxwell, 2003). Third, as discussed earlier, only three waves of data were simulated; the presence of more than three waves creates additional modeling complexities and both power and parameter estimation accuracy may vary significantly from the three-wave model. Fourth, only unreliability in the mediator was varied in the present study, the simplifying assumption being that the X and Y variables were measured without error. The impact of measurement error when

all three variables in the autoregressive mediation model are imperfectly measured can lead to a confusing pattern of underestimation and overestimation, and has led some to conclude that such impacts may be “impossible to predict beforehand” (Little et al., 2007, p. 361). Fifth, the present study only examined cross-wave, within-construct error covariance; other types of shared method variance are possible, and depending on the measurement of X, M, and Y, these effects can be extracted using structural equation modeling (Cole & Maxwell, 2003).

Sixth, we assumed the case where all three variables are measured at all occasions. It is possible to use an autoregressive mediation model where X is a time-invariant variable that reflects exposure to an intervention (e.g., see Roth & MacKinnon, 2012). While other studies of mediation from a cross-sectional perspective have shown that the categorical/continuous distinction for the X variable have little impact on study outcomes when other conditions are equivalent (MacKinnon et al., 2002), because of the additional paths found in an autoregressive mediation model, these findings may not generalize. Seventh, the indicators of the latent variable M were assumed continuous in the present study. The modeling of ordered categorical data (e.g., items on a Likert scale) has received considerable attention in the structural equation modeling literature (Finney & DiStefano, 2006). Future studies should consider the impact of the inclusion of ordinal-level data as indicators on parameter estimation and power in the autoregressive mediation model. Finally, we generated data such that the X and Y variables as well as the latent variable M were continuous. Future research should examine the application of the autoregressive mediation model when all or some of these variables are categorical, including binary, nominal, ordinal, or count variables.

Conclusion

The autoregressive mediation model provides researchers with a valuable tool for investigating causal processes to explain how or why variables exert their influence on other variables. The present study demonstrates that failure to account for measurement error and shared method variance can have a significant impact on parameter estimation in this model. Both overestimation and underestimation of paths of interest can result, potentially leading to a misinterpretation of the mediation mechanism. Although the extraction of latent variables from multiple observed measures generally provides accurate estimates and also allows researchers to take into account method effects by allowing correlated measurement errors, latent variable models still require significant levels of composite reliability to achieve acceptable levels of power to detect the mediated effect.

References

- Anderson, S. E., & Williams, L. J. (1992). Assumptions about unmeasured variables with studies of reciprocal relationships: The case of employee attitudes. *Journal of Applied Psychology, 77*, 638-650.
- Audrain-McGovern, J., Rodriguez, D., & Kassel, J. D. (2009). Adolescent smoking and depression: Evidence for self-medication and peer smoking mediation. *Addiction, 104*, 1743-1756.
- 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.
- Bedeian, A. G., Day, D. V., & Kelloway, E. K. (1997). Correcting for measurement error attenuation in structural equation models: Some important reminders. *Educational and Psychological Measurement, 57*, 785-799.
- Brown, T. A. (2006). *Confirmatory Factor Analysis for Applied Research*. New York: The Guilford Press.

- 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. (2011). Accuracy of estimates and statistical power for testing mediation in latent growth curve modeling. *Structural Equation Modeling, 18*, 195-211.
- Cheung, G. W., & Lau, R. S. (2008). Testing mediation and suppression effects of latent variables: Bootstrapping with structural equation models. *Organizational Research Methods, 11*, 296-325.
- Cheung, M. W. L. (2007). Comparison of approaches to constructing confidence intervals for mediating effects using structural equation models. *Structural Equation Modeling, 14*, 227-246.
- Cheung, M. W. L. (2009). Comparison of methods for constructing confidence intervals of standardized indirect effects. *Behavior Research Methods, 41*, 425-438.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*, 2nd ed. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Cohen, J., & Cohen, P. (1983). *Applied multiple regression/correlation analysis for the behavioral sciences*, 2nd ed. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Cole, D. A., & Maxwell, S. E. (2003). Testing mediation models with longitudinal data: Questions and tips in the use of structural equation modeling. *Journal of Abnormal Psychology, 112*, 558-577.
- Cole, D. A., Martin, N. C., & Steiger, J. H. (2005). Empirical and conceptual problems with longitudinal trait-state models: Introducing a trait-state-occasion model. *Psychological Methods, 10*, 3-20.
- Cole, D. A., Ciesla, J. A., & Steiger, J. H. (2007). The insidious effects of failing to include design-driven correlated residuals in latent-variable covariance structure analysis. *Psychological Methods, 12*, 381-398.
- Collins, L. M., Graham, J. W., & Flaherty, B. P. (1998). An alternative framework for defining mediation. *Multivariate Behavioral Research, 33*, 295-312.
- Collins, L. M., & Graham, J. W. (2002). The effect of the timing and spacing of observations in longitudinal studies of tobacco and other drug use: temporal design considerations. *Drug and Alcohol Dependence, 68*, S85-S96.
- Curran, P. J., & Bollen, K. A. (2001). The best of both worlds: Combining autoregressive and latent curve models. In L. M. Collins & A. G. Sayer (Eds.), *New methods for the analysis of change* (pp. 107-135). Washington, DC: American Psychological Association.
- Eid, M., Schneider, C., & Schwenkmezger, P. (1999). Do you feel better or worse? The validity of perceived deviations of mood states from mood traits. *European Journal of Personality, 13*, 283-306.
- Ferrer, E., & McArdle, J. J. (2003). Alternative structural models for multivariate longitudinal data analysis. *Structural Equation Modeling, 10*, 493-524.
- Finney, S. J., & DiStefano C. (2006). Non-normal and categorical data in structural equation modeling. In G. R. Hancock & R. O. Mueller (Eds.), *Structural equation modeling: A second course* (pp. 269-314). Greenwich, CT: Information Age Publishing.

- Flora, D. B., Khoo, S., & Chassin, L. (2007). Moderating effects of a risk factor: Modeling longitudinal moderated mediation in the development of adolescent heavy drinking. In T. D. Little, J. A. Bovaird, N. A. Card (Eds.), *Modeling contextual effects in longitudinal studies* (pp. 231-254). Mahwah, NJ: Lawrence Erlbaum Associates Publishers.
- Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research*, 18, 39-50.
- Geiser, C. G., Eid, M., Nussbeck, F. W., Courvoisier, D. S., & Cole, D. A. (2010). Analyzing true change in longitudinal multitrait-multimethod studies: Application of a multimethod change model to depression and anxiety in children. *Developmental Psychological*, 46, 29-45.
- 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.
- Hoyle, R. H., & Kenny, D. A. (1999). Sample size, reliability, and tests of statistical mediation. In R. H. Hoyle (Ed.), *Statistical strategies for small sample research* (pp. 195-222). Thousand Oaks, CA: Sage.
- Kaplan, D. (1988). The impact of specification error on the estimation, testing, and improvement of structural equation models. *Multivariate Behavioral Research*, 23, 69-86.
- Kazdin, A. E., & Nock, M. K. (2003). Delineating mechanisms of change in child and adolescent therapy: methodological issues and research recommendations. *Journal of Child Psychology and Psychiatry*, 44, 1116-1129.
- Kenny, D. A., Korchmaros, J. D., & Bolger N. (2003). Lower level mediation in multilevel models. *Psychological Methods*, 8, 115-128.
- Kline, R. B. (2011). *Principles and practice of structural equation modeling*, 3rd ed. New York: The Guilford Press.
- Little, T. D., Preacher, K. J., Selig, J. P., & Card, N. A. (2007). New developments in latent variable panel analyses of longitudinal data. *International Journal of Behavioral Development*, 31, 357-365.
- Littlefield, A. K., Sher, K. J., & Wood, P. K. (2010). Do changes in drinking motives mediate the relation between personality change and “maturing out” of problem drinking? *Journal of Abnormal Psychology*, 119, 93-105.
- Liu, L. C., Flay, B. R., & Aban Aya Investigators. (2009). Evaluating mediation in longitudinal multivariate data: Mediation effects for the Aban Aya youth project drug prevention program. *Prevention Science*, 10, 197-207.
- MacKinnon, D. P., Lockwood, C. M., Hoffman, J. M., West, S. G., & Sheets, V. (2002). A comparison of methods to test the significance of the mediated effect. *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., 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. (2008a). *Introduction to statistical mediation analysis*. New York: Lawrence Erlbaum Associates.
- MacKinnon, D. P. (2008b). Longitudinal mediation analysis. In D. P. MacKinnon, *Introduction to statistical mediation analysis* (pp. 193-236). New York: Lawrence Erlbaum Associates.
- Marsh, H. W. (1993). Stability of individual differences in multiwave panel studies: Comparison of simplex models and one-factor models. *Journal of Educational Measurement*, 30, 157-183.
- Maxwell, S. E., & Cole, D. A. (2007). Bias in cross-sectional analyses of longitudinal mediation. *Psychological Methods*, 12, 23-44.
- McArdle, J. J. (2009). Latent variable modeling of differences and changes with longitudinal data. *Annual Review of Psychology*, 60, 577-605.
- Negriff, S., Ji, J., & Trickett, P. (2011). Exposure to peer delinquency as a mediator between self-report pubertal timing and delinquency: a longitudinal study of mediation. *Development and Psychopathology*, 23, 293-304.
- Paxton, P., Curran, P. J., Bollen, K. A., Kirby, J., & Chen F. (2001). Monte Carlo experiments: Design and implementation. *Structural Equation Modeling*, 8, 287-312.
- Pössel, P., & Thomas, S. (2011). Cognitive triad as mediator in the hopelessness model? A three-wave longitudinal study. *Journal of Clinical Psychology*, 67, 224-240.
- Raffalovich, L. E., & Bohrnstedt, G. W. (1987). Common, specific, and error variance components of factor models: Estimation with longitudinal data. *Sociological Methods & Research*, 15, 385-405.
- Raykov, T., & Penev, S. (2005). Estimation of reliability for multiple-component measuring instruments in test-retest designs. *British Journal of Mathematical and Statistical Psychology*, 58, 285-299.
- Roesch, S. C., Norman, G. J., Adams, M. A., Kerr, J., Sallis, J. F., Ryan, S., & ... Patrick, K. (2009). Latent growth curve modeling of adolescent physical activity: Testing parallel process and mediation models. *Journal of Health Psychology*, 14, 313-325.
- Roesch, S. C., Norman, G. J., Villodas, F., Sallis, J. F., & Patrick, K. (2010). Intervention-mediated effects for adult physical activity: A latent growth curve analysis. *Social Science & Medicine*, 71, 494-501.
- Roth, D. L., & MacKinnon, D. P. (2012). Mediation analysis with longitudinal data. In J. T. Newsom, R. N. Jones, & S. M. Hofer (Eds.), *Longitudinal data analysis: A practical guide for researchers in aging, health, and social sciences* (pp. 181-216). New York: Routledge.
- Selig, J. P., & Preacher, K. J. (2009). Mediation models for longitudinal data in developmental research. *Research in Human Development*, 6, 144-164.

- Shrout, P. E., & Bolger, N. (2002). Mediation in experimental and nonexperimental studies: New procedures and recommendations. *Psychological Methods*, 7, 422-445.
- Sobel, M. E. (1982). Asymptotic confidence intervals for indirect effects in structural equation models. *Sociological Methodology*, 13, 290-312.
- Stephenson, M. T., & Holbert, R. L. (2003). A Monte Carlo simulation of observable versus latent variable structural equation modeling techniques. *Communication Research*, 30, 332-354.
- Taylor, A. B., MacKinnon, D. P., Tein J. Y. (2008). Tests of the three-path mediated effect. *Organizational Research Methods*, 11, 241-269.
- von Soest, T., & Hagtvet, K. A. (2011). Mediation analysis in a latent growth curve modeling framework. *Structural Equation Modeling*, 18, 289-314.

Table 1. Mean relative bias of the mediated effect estimates ($a*b$) across simulation conditions

		Model for the mediator and sample size											
Reliability of the mediator	Indicator error correlation	Latent variable model with indicator error covariances specified				Latent variable model with indicator error covariances unspecified				Composite variable model			
		$n = 100$	$n = 200$	$n = 500$	$n = 1000$	$n = 100$	$n = 200$	$n = 500$	$n = 1000$	$n = 100$	$n = 200$	$n = 500$	$n = 1000$
Low	None	0.037	0.005	0.006	0.003	0.032	0.003	0.005	0.003	-0.332	-0.345	-0.336	-0.338
	Small	0.034	0.031	0.001	-0.003	-0.091	-0.069	-0.096	-0.098	-0.369	-0.356	-0.362	-0.364
	Medium	0.050	0.017	0.009	0.004	-0.255	-0.285	-0.270	-0.272	-0.406	-0.412	-0.409	-0.409
Moderate	None	-0.022	-0.006	0.004	0.002	-0.022	-0.005	0.004	0.002	-0.226	-0.208	-0.198	-0.201
	Small	0.001	-0.005	-0.017	0.001	-0.046	-0.053	-0.062	-0.044	-0.231	-0.225	-0.231	-0.218
	Medium	0.010	-0.002	0.010	0.013	-0.134	-0.143	-0.128	-0.123	-0.261	-0.263	-0.252	-0.249
High	None	-0.004	0.006	0.009	0.002	-0.001	0.006	0.009	0.002	-0.078	-0.069	-0.067	-0.073
	Small	-0.009	-0.009	0.011	-0.008	-0.025	-0.023	-0.003	-0.021	-0.099	-0.093	-0.075	-0.091
	Medium	0.001	0.011	-0.007	-0.002	-0.042	-0.034	-0.042	-0.043	-0.109	-0.096	-0.106	-0.106

Note: Entries are mean relative bias, collapsing across factors for which results were similar (i.e., degree of stability of the mediator (or size of path m) and size of the mediated effect, excluding the zero mediated effect, as relative bias cannot be calculated under conditions where the true value was zero). Although sample size or interactions with this factor did not account for nontrivial variance, it is nevertheless tabled above. Reliability of the mediator was defined as low = 0.60, medium = 0.75, and high = 0.90 and indicator error correlation (i.e., shared method variance) was defined as none = 0, small = 0.10, and medium = 0.30.

Table 2. Mean relative bias of the $M_1 \rightarrow M_2$ path estimates (path m) across simulation conditions

		Model for the mediator and sample size											
Reliability of the mediator	Indicator error correlation	Latent variable model with indicator error covariances specified				Latent variable model with indicator error covariances unspecified				Composite variable model			
		$n = 100$	$n = 200$	$n = 500$	$n = 1000$	$n = 100$	$n = 200$	$n = 500$	$n = 1000$	$n = 100$	$n = 200$	$n = 500$	$n = 1000$
Low	None	0.051	0.036	0.013	0.004	0.054	0.038	0.011	0.003	-0.411	-0.413	-0.413	-0.414
	Small	0.063	0.028	0.010	0.004	0.191	0.160	0.138	0.129	-0.315	-0.314	-0.315	-0.315
	Medium	0.044	0.013	0.005	0.004	0.436	0.386	0.367	0.364	-0.113	-0.118	-0.120	-0.119
Moderate	None	0.021	0.006	0.006	0.002	0.022	0.005	0.006	0.002	-0.256	-0.262	-0.259	-0.261
	Small	0.022	0.013	0.002	-0.001	0.094	0.086	0.074	0.070	-0.196	-0.194	-0.200	-0.199
	Medium	0.006	0.009	0.002	0.002	0.224	0.220	0.207	0.206	-0.077	-0.074	-0.076	-0.076
High	None	0.004	0.005	-0.001	0.001	0.004	0.005	-0.001	0.001	-0.108	-0.102	-0.107	-0.105
	Small	0.005	0.005	-0.0001	0.001	0.032	0.032	0.026	0.026	-0.080	-0.078	-0.081	-0.080
	Medium	0.002	0.001	-0.001	0.001	0.083	0.079	0.077	0.078	-0.030	-0.031	-0.031	-0.030

Note: Entries are mean relative bias, collapsing across factors for which results were similar (i.e., degree of stability of the mediator (or size of path m) and size of the mediated effect). Although sample size or interactions with this factor did not account for nontrivial variance, it is nevertheless tabled above. Reliability of the mediator was defined as low = 0.60, medium = 0.75, and high = 0.90 and indicator error correlation (i.e., shared method variance) was defined as none = 0, small = 0.10, and medium = 0.30.

Table 3. Mean relative bias of the $X_1 \rightarrow M_2$ path estimates (path a) across simulation conditions

		Model for the mediator and sample size											
Reliability of the mediator	Indicator error correlation	Latent variable model with indicator error covariances specified				Latent variable model with indicator error covariances unspecified				Composite variable model			
		$n = 100$	$n = 200$	$n = 500$	$n = 1000$	$n = 100$	$n = 200$	$n = 500$	$n = 1000$	$n = 100$	$n = 200$	$n = 500$	$n = 1000$
Low	None	0.011	-0.001	0.005	-0.0002	0.014	-0.003	0.005	-0.0002	0.220	0.191	0.198	0.191
	Small	-0.026	-0.0001	-0.002	-0.008	-0.095	-0.061	-0.063	-0.068	0.142	0.143	0.153	0.141
	Medium	0.005	0.001	0.001	-0.004	-0.178	-0.185	-0.178	-0.180	0.068	0.060	0.063	0.056
Moderate	None	-0.030	0.001	0.003	0.006	-0.030	0.001	0.002	0.006	0.098	0.124	0.123	0.125
	Small	0.005	-0.004	-0.013	-0.006	-0.028	-0.038	-0.047	-0.039	0.102	0.090	0.081	0.088
	Medium	-0.023	0.001	-0.001	0.006	-0.125	-0.099	0.101	-0.091	0.021	0.041	0.036	0.044
High	None	-0.007	0.009	-0.002	0.004	-0.006	0.009	-0.003	0.004	0.042	0.058	0.046	0.053
	Small	-0.017	-0.004	0.007	-0.003	-0.030	-0.016	0.005	-0.014	0.022	0.035	0.044	0.035
	Medium	0.008	-0.001	0.003	-0.005	-0.030	-0.038	-0.033	-0.040	0.027	0.014	0.018	0.011

Note: Entries are mean relative bias, collapsing across factors for which results were similar (i.e., degree of stability of the mediator (or size of path m) and size of the mediated effect, excluding the zero mediated effect, as relative bias cannot be calculated under conditions where the true value was zero). Although sample size or interactions with this factor did not account for nontrivial variance, it is nevertheless tabled above. Reliability of the mediator was defined as low = 0.60, medium = 0.75, and high = 0.90 and indicator error correlation (i.e., shared method variance) was defined as none = 0, small = 0.10, and medium = 0.30.

Table 4. Mean relative bias of the $M_2 \rightarrow Y_3$ path estimates (path b) across simulation conditions

		Model for the mediator and sample size											
Reliability of the mediator	Indicator error correlation	Latent variable model with indicator error covariances specified				Latent variable model with indicator error covariances unspecified				Composite variable model			
		$n = 100$	$n = 200$	$n = 500$	$n = 1000$	$n = 100$	$n = 200$	$n = 500$	$n = 1000$	$n = 100$	$n = 200$	$n = 500$	$n = 1000$
Low	None	0.068	0.026	0.004	0.002	0.058	0.026	0.003	0.002	-0.447	-0.446	-0.446	-0.445
	Small	0.103	0.053	0.010	0.006	0.040	0.007	-0.028	-0.032	-0.437	-0.434	-0.445	-0.443
	Medium	0.082	0.031	0.016	0.012	-0.073	-0.104	-0.105	-0.108	-0.433	-0.442	-0.442	-0.439
Moderate	None	0.015	0.003	0.004	-0.003	0.013	0.003	0.004	-0.003	-0.293	-0.290	-0.285	-0.289
	Small	0.012	0.011	0.002	0.007	-0.004	-0.004	-0.011	-0.006	-0.297	-0.285	-0.287	-0.282
	Medium	0.040	-0.001	0.012	0.008	-0.004	-0.046	-0.029	-0.034	-0.276	-0.293	-0.279	-0.281
High	None	-0.005	0.004	0.008	-0.002	-0.004	0.004	0.008	-0.002	-0.125	-0.116	-0.112	-0.120
	Small	0.007	-0.001	0.007	-0.004	0.004	-0.004	0.005	-0.005	-0.119	-0.123	-0.112	-0.121
	Medium	0.001	0.022	-0.001	0.002	-0.005	0.016	-0.007	-0.003	-0.126	-0.101	-0.120	-0.117

Note: Entries are mean relative bias, collapsing across factors for which results were similar (i.e., degree of stability of the mediator (or size of path m) and size of the mediated effect, excluding the zero mediated effect, as relative bias cannot be calculated under conditions where the true value was zero). Although sample size or interactions with this factor did not account for nontrivial variance, it is nevertheless tabled above. Reliability of the mediator was defined as low = 0.60, medium = 0.75, and high = 0.90 and indicator error correlation (i.e., shared method variance) was defined as none = 0, small = 0.10, and medium = 0.30.

Table 5. Mean bias of the $X_1 \rightarrow Y_2$ path estimates (path c_1') (i.e., first direct effect) across simulation conditions

		Model for the mediator and sample size											
Reliability of the mediator	Effect size of mediated effect	Latent variable model with indicator error covariances specified				Latent variable model with indicator error covariances unspecified				Composite variable model			
		$n = 100$	$n = 200$	$n = 500$	$n = 1000$	$n = 100$	$n = 200$	$n = 500$	$n = 1000$	$n = 100$	$n = 200$	$n = 500$	$n = 1000$
Low	Zero	-0.001	-0.001	-0.001	0.0002	-0.001	-0.001	-0.001	0.0001	-0.001	-0.001	-0.001	0.0001
	Small	0.001	-0.001	0.0003	-0.001	0.002	-0.001	0.001	-0.0003	0.006	0.003	0.005	0.004
	Medium	-0.001	-0.001	-0.0000	-0.001	0.002	0.001	0.002	0.001	0.027	0.026	0.027	0.026
	Large	-0.005	-0.001	-0.001	-0.001	-0.0002	0.002	0.002	0.003	0.048	0.049	0.048	0.048
Moderate	Zero	-0.001	-0.001	0.001	-0.001	-0.001	-0.0004	0.001	-0.0004	-0.001	-0.001	0.001	-0.001
	Small	-0.001	0.001	-0.0003	-0.0004	-0.001	0.001	-0.0002	-0.0003	0.001	0.004	0.002	0.002
	Medium	0.0001	-0.001	-0.001	0.001	0.001	-0.0002	-0.0002	0.001	0.017	0.016	0.016	0.018
	Large	0.001	-0.001	0.0001	-0.0001	0.002	-0.0003	0.001	0.001	0.034	0.031	0.032	0.032
High	Zero	-0.0002	0.0004	0.001	0.001	-0.0002	0.0004	0.001	0.001	-0.0002	0.001	0.001	0.001
	Small	-0.004	-0.003	-0.001	0.0004	-0.004	-0.003	-0.001	0.0004	-0.002	-0.002	0.001	0.001
	Medium	-0.001	-0.001	0.0003	-0.0001	-0.001	-0.0004	0.0003	-0.0001	0.006	0.006	0.007	0.007
	Large	-0.002	-0.0004	-0.002	-0.0004	-0.002	-0.0003	-0.002	-0.0002	0.012	0.013	0.011	0.013

Note: Since the population value is 0, entries are mean bias not relative bias, collapsing across factors for which results were similar (i.e., degree of stability of the mediator (or size of path m) and level of indicator error correlation). Although sample size or interactions with this factor did not account for nontrivial variance, it is nevertheless tabled above. Reliability of the mediator was defined as low = 0.60, medium = 0.75, and high = 0.90 and effect size of the mediated effect was defined according to values on the f^2 metric for the incremental effect of including the a and b paths in the autoregressive model over and above the y path as described by Cohen (1988).

Table 6. Mean bias of the $X_1 \rightarrow Y_3$ path estimates (path c_2') (i.e., second direct effect) across simulation conditions

		Model for the mediator and sample size											
Reliability of the mediator	Effect size of mediated effect	Latent variable model with indicator error covariances specified				Latent variable model with indicator error covariances unspecified				Composite variable model			
		$n = 100$	$n = 200$	$n = 500$	$n = 1000$	$n = 100$	$n = 200$	$n = 500$	$n = 1000$	$n = 100$	$n = 200$	$n = 500$	$n = 1000$
Low	Zero	0.003	0.0002	-0.0001	0.001	0.003	0.0002	-0.0001	0.001	0.003	0.0003	-0.0001	0.001
	Small	-0.001	-0.0001	0.001	-0.001	-0.0002	0.0007	0.002	-0.001	0.006	0.006	0.007	0.005
	Medium	-0.004	-0.002	-0.0003	-0.001	0.002	0.003	0.005	0.005	0.037	0.037	0.039	0.038
	Large	-0.009	-0.002	-0.001	-0.002	0.006	0.010	0.011	0.010	0.076	0.075	0.076	0.075
Moderate	Zero	-0.002	0.0001	0.001	-0.001	-0.002	0.0001	0.001	-0.001	-0.002	0.0000	0.001	-0.001
	Small	-0.002	0.0001	-0.0001	-0.0003	-0.001	0.0003	0.0002	-0.0001	0.002	0.004	0.004	0.003
	Medium	-0.001	-0.002	0.001	0.0004	0.001	-0.001	0.002	0.002	0.024	0.023	0.025	0.025
	Large	-0.002	-0.0004	-0.0001	0.0001	0.003	0.004	0.004	0.005	0.050	0.050	0.050	0.050
High	Zero	-0.0002	-0.001	0.001	-0.001	-0.0001	-0.001	0.001	-0.001	-0.0001	-0.001	0.001	-0.001
	Small	0.001	-0.001	0.0001	-0.0003	0.001	-0.001	0.0001	-0.0003	0.002	0.001	0.002	0.001
	Medium	-0.001	0.001	-0.002	0.001	-0.001	0.001	-0.001	0.001	0.001	0.011	0.009	0.011
	Large	-0.002	0.0002	-0.0004	-0.0003	-0.001	0.001	0.0003	0.0004	0.020	0.022	0.021	0.021

Note: Since the population value is 0, entries are mean bias not relative bias, collapsing across factors for which results were similar (i.e., degree of stability of the mediator (or size of path m) and level of indicator error correlation). Although sample size or interactions with this factor did not account for nontrivial variance, it is nevertheless tabled above. Reliability of the mediator was defined as low = 0.60, medium = 0.75, and high = 0.90 and effect size of the mediated effect was defined according to values on the f^2 metric for the incremental effect of including the a and b paths in the autoregressive model over and above the y path as described by Cohen (1988).

Table 7. Mean relative bias of standard error of the mediated effect ($a*b$) across simulation conditions

		Model for the mediator and sample size											
Reliability of the mediator	Effect size of mediated effect	Latent variable model with indicator error covariances specified				Latent variable model with indicator error covariances unspecified				Composite variable model			
		$n = 100$	$n = 200$	$n = 500$	$n = 1000$	$n = 100$	$n = 200$	$n = 500$	$n = 1000$	$n = 100$	$n = 200$	$n = 500$	$n = 1000$
Low	Zero	0.172	0.165	0.226	0.252	0.193	0.178	0.228	0.239	0.207	0.200	0.255	0.273
	Small	-0.019	0.001	0.003	0.002	0.036	0.021	0.011	0.006	0.035	0.019	0.006	0.006
	Medium	-0.093	-0.033	-0.009	-0.008	-0.052	-0.018	-0.003	-0.006	-0.002	0.003	0.006	0.001
	Large	-0.140	-0.004	-0.005	-0.011	-0.070	0.015	0.001	-0.007	-0.008	0.009	0.010	0.001
Moderate	Zero	0.183	0.224	0.231	0.245	0.195	0.229	0.234	0.239	0.206	0.239	0.224	0.253
	Small	0.010	0.003	0.004	0.009	0.026	0.014	0.006	0.011	0.037	0.012	-0.002	0.015
	Medium	-0.045	-0.019	-0.022	-0.002	-0.036	-0.020	-0.019	0.001	-0.015	-0.013	-0.009	-0.0002
	Large	-0.071	-0.017	-0.007	0.007	-0.061	-0.012	-0.0004	0.008	-0.031	-0.0001	-0.003	0.015
High	Zero	0.221	0.206	0.247	0.247	0.224	0.208	0.247	0.249	0.235	0.211	0.250	0.245
	Small	-0.002	0.021	-0.005	0.011	-0.003	0.022	-0.004	0.012	0.005	0.022	-0.003	0.011
	Medium	-0.010	-0.023	0.006	-0.002	-0.004	-0.023	0.007	-0.001	0.0004	-0.019	0.006	-0.002
	Large	-0.029	-0.007	-0.034	0.0001	-0.026	-0.009	-0.032	0.0000	-0.017	-0.002	-0.035	0.002

Note: Entries are mean relative bias, collapsing across factors for which results were similar (i.e., degree of stability of the mediator (or size of path m) and level of indicator error correlation). Standard error of the mediated effect was obtained using Sobel's (1982) method. Reliability of the mediator was defined as low = 0.60, medium = 0.75, and high = 0.90 and effect size of the mediated effect was defined according to values on the f^2 metric for the incremental effect of including the a and b paths in the autoregressive model over and above the y path as described by Cohen (1988).

Table 8. Estimated Type I error rates for the mediated effect across simulation conditions

Method for testing mediated effect	Model for the mediator and sample size											
	Latent variable model with indicator error covariances specified				Latent variable model with indicator error covariances unspecified				Composite variable model			
	<i>n</i> = 100	<i>n</i> = 200	<i>n</i> = 500	<i>n</i> = 1000	<i>n</i> = 100	<i>n</i> = 200	<i>n</i> = 500	<i>n</i> = 1000	<i>n</i> = 100	<i>n</i> = 200	<i>n</i> = 500	<i>n</i> = 1000
Sobel test	0.0003	0.0002	0.0001	0.0001	0.0002	0.0002	0.0001	0.0001	0.0002	0.0003	0.0002	0.0001
Joint significance test	0.0029	0.0041	0.0022	0.0026	0.0032	0.0039	0.0023	0.0024	0.0044	0.0039	0.0022	0.0026

Note: Entries are mean Type I error rates for the mediated effect, collapsing across factors for which results were similar (i.e., degree of stability of the mediator (or size of path *m*), reliability of the mediator, and level of indicator error correlation).

Table 9. Empirical power levels across simulation conditions for the test of the mediated effect

		Model for the mediator and sample size											
Reliability of the mediator	Effect size of mediated effect	Latent variable model with indicator error covariances specified				Latent variable model with indicator error covariances unspecified				Composite variable model			
		<i>n</i> = 100	<i>n</i> = 200	<i>n</i> = 500	<i>n</i> = 1000	<i>n</i> = 100	<i>n</i> = 200	<i>n</i> = 500	<i>n</i> = 1000	<i>n</i> = 100	<i>n</i> = 200	<i>n</i> = 500	<i>n</i> = 1000
Sobel test													
Low	Small	0.003	0.009	0.074	0.344	0.003	0.006	0.057	0.283	0.004	0.010	0.096	0.442
	Medium	0.086	0.422	0.971	1.000	0.069	0.354	0.942	0.999	0.167	0.578	0.993	1.000
	Large	0.262	0.809	0.999	1.000	0.249	0.746	0.997	1.000	0.518	0.934	1.000	1.000
Moderate	Small	0.005	0.016	0.135	0.554	0.003	0.013	0.115	0.519	0.005	0.019	0.165	0.616
	Medium	0.203	0.681	0.998	1.000	0.179	0.640	0.994	1.000	0.270	0.764	0.999	1.000
	Large	0.570	0.971	1.000	1.000	0.537	0.957	1.000	1.000	0.702	0.983	1.000	1.000
High	Small	0.005	0.022	0.217	0.712	0.004	0.021	0.209	0.700	0.005	0.023	0.230	0.739
	Medium	0.334	0.849	1.000	1.000	0.319	0.838	1.000	1.000	0.367	0.868	1.000	1.000
	Large	0.807	0.997	1.000	1.000	0.789	0.996	1.000	1.000	0.843	0.998	1.000	1.000
Joint significance test													
Low	Small	0.011	0.043	0.221	0.585	0.012	0.039	0.181	0.513	0.028	0.070	0.276	0.667
	Medium	0.221	0.634	0.986	1.000	0.196	0.562	0.966	1.000	0.358	0.761	0.995	1.000
	Large	0.475	0.900	1.000	1.000	0.442	0.848	0.998	1.000	0.708	0.963	1.000	1.000
Moderate	Small	0.027	0.072	0.328	0.750	0.026	0.064	0.307	0.712	0.034	0.089	0.369	0.795
	Medium	0.398	0.830	0.999	1.000	0.371	0.793	0.997	1.000	0.496	0.875	1.000	1.000
	Large	0.746	0.990	1.000	1.000	0.713	0.981	1.000	1.000	0.834	0.991	1.000	1.000
High	Small	0.037	0.105	0.454	0.851	0.035	0.100	0.445	0.841	0.041	0.117	0.475	0.864
	Medium	0.552	0.915	1.000	1.000	0.547	0.909	1.000	1.000	0.581	0.927	1.000	1.000
	Large	0.896	0.999	1.000	1.000	0.892	0.999	1.000	1.000	0.916	0.999	1.000	1.000

Note: Entries are mean power levels, collapsing across factors for which results were similar (i.e., degree of stability of the mediator (or size of path *m*) and level of indicator error correlation). Reliability of the mediator was defined as low = 0.60, medium = 0.75, and high = 0.90 and effect size of the mediated effect was defined according to values on the f^2 metric for the incremental effect of including the *a* and *b* paths in the autoregressive model over and above the *y* path as described by Cohen (1988).

Table 10. Estimated Type I error rates for the direct effects across simulation conditions levels

		Model for the mediator and sample size											
Reliability of the mediator	Effect size of mediated effect	Latent variable model with indicator error covariances specified				Latent variable model with indicator error covariances unspecified				Composite variable model			
		<i>n</i> = 100	<i>n</i> = 200	<i>n</i> = 500	<i>n</i> = 1000	<i>n</i> = 100	<i>n</i> = 200	<i>n</i> = 500	<i>n</i> = 1000	<i>n</i> = 100	<i>n</i> = 200	<i>n</i> = 500	<i>n</i> = 1000
$X_1 \rightarrow Y_2$ path (path c_1')													
Low	Zero	0.054	0.060	0.049	0.046	0.052	0.060	0.049	0.046	0.055	0.061	0.049	0.047
	Small	0.058	0.048	0.055	0.047	0.059	0.047	0.054	0.047	0.057	0.047	0.056	0.046
	Medium	0.052	0.046	0.056	0.060	0.051	0.045	0.056	0.059	0.072	0.073	0.116	0.164
	Large	0.056	0.051	0.044	0.057	0.059	0.051	0.047	0.055	0.107	0.151	0.289	0.487
Moderate	Zero	0.056	0.050	0.047	0.045	0.056	0.049	0.047	0.045	0.056	0.052	0.047	0.045
	Small	0.051	0.059	0.055	0.046	0.052	0.059	0.055	0.046	0.050	0.058	0.056	0.050
	Medium	0.055	0.051	0.050	0.049	0.054	0.048	0.050	0.050	0.060	0.056	0.077	0.098
	Large	0.055	0.055	0.051	0.047	0.055	0.056	0.054	0.048	0.081	0.097	0.153	0.253
High	Zero	0.054	0.053	0.052	0.053	0.054	0.053	0.052	0.053	0.054	0.053	0.053	0.053
	Small	0.058	0.051	0.055	0.053	0.058	0.050	0.055	0.052	0.058	0.052	0.055	0.052
	Medium	0.046	0.056	0.043	0.047	0.046	0.055	0.043	0.048	0.048	0.056	0.054	0.052
	Large	0.060	0.057	0.057	0.050	0.059	0.056	0.058	0.051	0.060	0.061	0.072	0.079
$X_1 \rightarrow Y_3$ path (path c_2')													
Low	Zero	0.062	0.055	0.052	0.046	0.062	0.056	0.052	0.046	0.061	0.055	0.053	0.047
	Small	0.050	0.055	0.053	0.048	0.049	0.054	0.054	0.047	0.050	0.059	0.058	0.056
	Medium	0.062	0.046	0.047	0.045	0.059	0.048	0.051	0.052	0.088	0.096	0.179	0.298
	Large	0.050	0.049	0.051	0.051	0.054	0.056	0.069	0.077	0.170	0.265	0.551	0.811
Moderate	Zero	0.059	0.056	0.051	0.048	0.058	0.056	0.051	0.048	0.059	0.055	0.051	0.049
	Small	0.050	0.051	0.052	0.055	0.051	0.052	0.052	0.054	0.051	0.051	0.052	0.060
	Medium	0.052	0.050	0.051	0.050	0.053	0.050	0.051	0.051	0.060	0.070	0.102	0.160
	Large	0.051	0.053	0.053	0.054	0.053	0.053	0.058	0.059	0.093	0.147	0.278	0.477
High	Zero	0.057	0.056	0.049	0.050	0.058	0.056	0.049	0.050	0.058	0.056	0.049	0.050
	Small	0.068	0.054	0.050	0.051	0.068	0.054	0.050	0.051	0.068	0.054	0.049	0.053
	Medium	0.059	0.052	0.050	0.052	0.059	0.050	0.049	0.052	0.060	0.056	0.059	0.073
	Large	0.057	0.052	0.057	0.049	0.057	0.052	0.057	0.049	0.062	0.072	0.093	0.131

Note: Entries are mean Type I error rates for the two specified direct effects, collapsing across factors for which results were similar (i.e., degree of stability of the mediator (or size of path m) and level of indicator error correlation). Reliability of the mediator was defined as low = 0.60, medium = 0.75, and high = 0.90 and effect size of the mediated effect was defined according to values on the f^2 metric for the incremental effect of including the a and b paths in the autoregressive model over and above the y path as described by Cohen (1988).

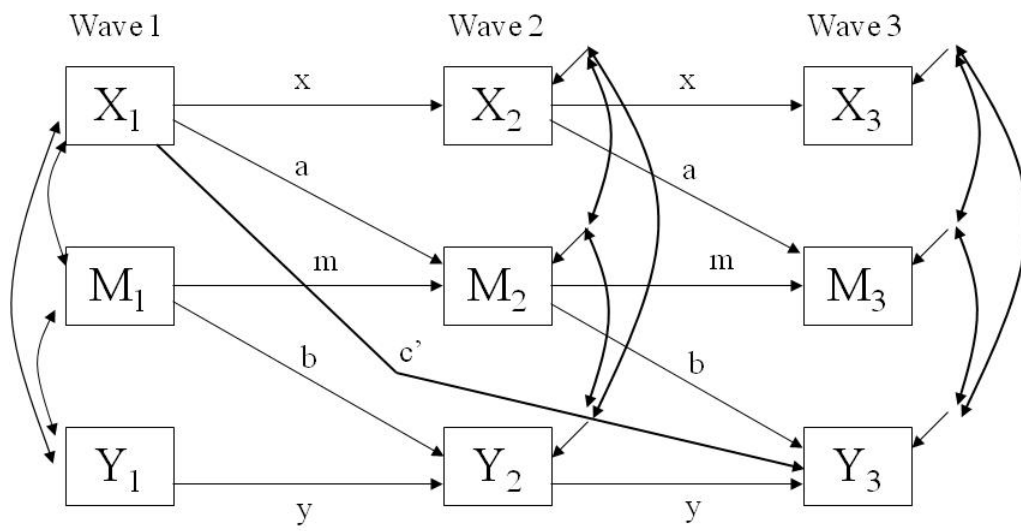


Figure 1. Basic three-wave autoregressive mediation model.

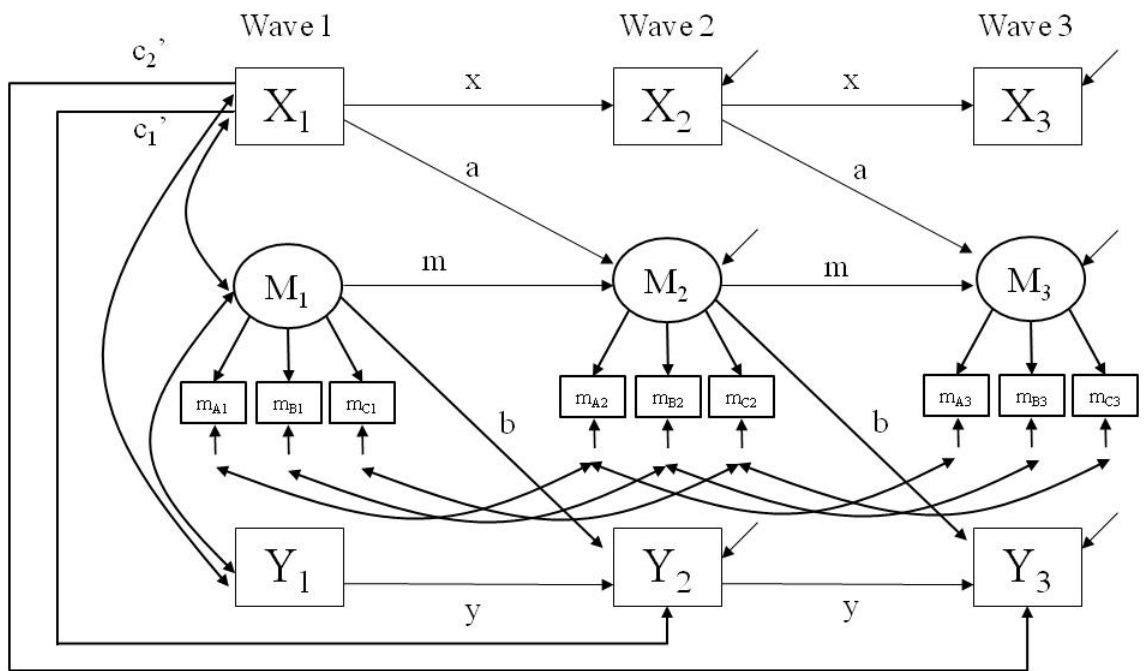


Figure 2. Basic three-wave autoregressive mediation model with a latent mediator variable.

Appendix A – Determining effect sizes for the mediated effect

Under a set of assumptions consistent with the current study (i.e., complete mediation, path coefficients a , b , x , m , and y are invariant over time, and a system at equilibrium), Maxwell & Cole (2007) show that zero-order correlations among X, M, and Y at each measurement occasion, t , (i.e., the cross-sectional correlations) are given by the following equations:

$$\rho_{X,M_t} = \frac{ax}{1-mx} \quad (A1)$$

$$\rho_{X,Y_t} = \frac{abx^2}{(1-mx)(1-xy)} \quad (A2)$$

$$\rho_{M_t,Y_t} = \frac{bm}{(1-my)} + \frac{a^2bx}{(1-mx)(1-my)(1-xy)} \quad (A3)$$

Furthermore, the variance of Y_3 is a function of the relationship of X_1 to Y_3 , M_2 to Y_3 , Y_2 to Y_3 , a number of covariances, and its residual variance. Note that the X_1 to Y_3 path is zero in the population and all population values were selected such that all X and Y observed variables and the M latent variables were standardized such that their standard deviations equal 1. In addition, the system was assumed to be at equilibrium such that the correlations among X, M, and Y are the same at all three occasions. Given these simplifying assumptions and using the notation in Figure 2:

$$Var(Y_3) = b^2 + y^2 + 2ba\rho_{XM}by + 2ba\rho_{XY}yy + 2bmby + 2bm\rho_{MY}yy + Var(\varepsilon_{Y_3}) \quad (A4)$$

Where $Var(Y_3)$ is the variance of Y at wave 3. Recall, that $y = 0.51$ and $x = 0.71$ were fixed values used in all simulation conditions. For the set of simulation conditions where $m = 0.36$ and $a = b = 0.34$, equation A1 yields a value of 0.324, equation A2 yields a value of 0.123, and equation A3 yields a value of 0.222. Setting the variance of Y_3 to 1, substituting known values of the path coefficients into equation A4, and solving for the residual variance of Y_3 yields a value of 0.547 (incidentally, at equilibrium, this is also the residual variance of Y_2 given the same parameter values). In *Mplus*, the residual variance of Y_3 was fixed at this value when generating data under these conditions. Given that the variance of Y_3 is 1, the amount of explained variance in Y_3 under these conditions is 45.3%. When $a = b = 0$ (i.e., no mediated effect), the amount of explained variance in Y_3 is 26% (all from the y path). Using Cohen's f^2 for incremental effect:

$$f^2 = \frac{\Delta R^2}{1 - R_{\text{Full Model}}^2} \quad (\text{A5})$$

yields a value of 0.35 under these conditions, which Cohen (1988) defines as a large effect. Similar calculations can be performed for $a = b = 0.1$ and $a = b = 0.25$, yielding Cohen's f^2 values classified as small and medium effect sizes (0.02 and 0.15), respectively. When $m = 0.51$, the effects are slightly larger, but still within Cohen's guidelines for small, medium, and large effects.

Appendix B – Mplus Source Code

!Mplus data generation code for true population model where reliability
!= 0.60, indicator error correlation=0.3, stability of mediator=0.36,
!size of the mediated effect = medium (a=b=0.25), and sample size=1000

MONTECARLO:

```
names = mA1 mB1 mC1 mA2 mB2 mC2 mA3 mB3 mC3 x2 x3 y2 y3 x1 y1;  
nobs = 1000;  
nreps = 600;  
seed = 6069894;  
repsave = all;  
save = cell_76rep*.dat;
```

MODEL MONTECARLO:

```
m1 by mA1@1 mB1@1 mC1@1;  
m2 by mA2@1 mB2@1 mC2@1;  
m3 by mA3@1 mB3@1 mC3@1;  
mA1@2;  
mB1@2;  
mC1@2;  
mA2@2;  
mB2@2;  
mC2@2;  
mA3@2;  
mB3@2;  
mC3@2;  
x1@1;  
x1 with m1@0.2384;  
x1 with y1@0.0663;  
m1 with y1@0.1389;  
x2 on x1@0.71;  
x3 on x2@0.71;  
x2@0.4959;  
x3@0.4959;  
m1@1;  
m2 on m1@0.36;  
m2 on x1@0.25;  
m3 on m2@0.36 x2@0.25;  
M2@0.7650;  
M3@0.7650;  
y1@1;  
y2 on y1@0.51 m1@0.25;  
y3 on y2@0.51;  
y3 on m2@0.25;  
y2@0.6420;  
y3@0.6420;  
mA1 with mA2@0.6;  
mB1 with mB2@0.6;  
mC1 with mC2@0.6;  
mA2 with mA3@0.6;  
mB2 with mB3@0.6;  
mC2 with mC3@0.6;
```

!Mplus code that reads in the data created from the first syntax and
!fits the latent variable model with indicator error covariances
!specified (i.e., freely estimated)

```
DATA:    FILE=cell_76replis.dat;
        TYPE = MONTECARLO;
VARIABLE: NAMES = mA1 mB1 mC1 mA2 mB2 mC2 mA3 mB3 mC3 x2 x3
          y2 y3 x1 y1;
        USEVARIABLES = ALL;
MODEL:
    m1 by mA1@1 mB1*1 mC1*1;
    m2 by mA2@1 mB2*1 mC2*1;
    m3 by mA3@1 mB3*1 mC3*1;
    mA1*2;
    mB1*2;
    mC1*2;
    mA2*2;
    mB2*2;
    mC2*2;
    mA3*2;
    mB3*2;
    mC3*2;
    x1*1;
    x1 with m1*0.2384;
    x1 with y1*0.0663;
    m1 with y1*0.1389;
    x3 with m3@0; !Fix this or Mplus will estimate by default
    m3 with y3@0; !Fix this or Mplus will estimate by default
    x3 with y3@0; !Fix this or Mplus will estimate by default
    x2 on x1*0.71;
    x3 on x2*0.71;
    x2*0.4959;
    x3*0.4959;
    m1*1;
    m2 on m1*0.36;
    m2 on x1*0.25 (a);
    m3 on m2*0.36 x2*0.25;
    M2*0.7650;
    M3*0.7650;
    y1*1;
    y2 on y1*0.51 m1*0.25 x1*0;
    y3 on y2*0.51;
    y3 on m2*0.25 (b);
    y3 on x1*0;
    y2*0.6420;
    y3*0.6420;
    mA1 with mA2*0.6;
    mB1 with mB2*0.6;
    mC1 with mC2*0.6;
    mA2 with mA3*0.6;
    mB2 with mB3*0.6;
    mC2 with mC3*0.6;
MODEL CONSTRAINT:
    NEW(med*0.0625);
    med=a*b;
OUTPUT: TECH9;
SAVEDATA: RESULTS=cell_76.txt;
```

!Mplus code that reads in the data created from the first syntax and
!fits the latent variable model with indicator error covariances
!unspecified (i.e., not estimated or fixed to zero)

```
DATA:  FILE=cell_76replis.dat;
      TYPE = MONTECARLO;
VARIABLE:  NAMES = mA1 mB1 mC1 mA2 mB2 mC2 mA3 mB3 mC3 x2 x3
            y2 y3 x1 y1;
          USEVARIABLES = ALL;
MODEL:
  m1 by mA1@1 mB1*1 mC1*1;
  m2 by mA2@1 mB2*1 mC2*1;
  m3 by mA3@1 mB3*1 mC3*1;
  mA1*2;
  mB1*2;
  mC1*2;
  mA2*2;
  mB2*2;
  mC2*2;
  mA3*2;
  mB3*2;
  mC3*2;
  x1*1;
  x1 with m1*0.2384;
  x1 with y1*0.0663;
  m1 with y1*0.1389;
  x3 with m3@0; !Fix this or Mplus will estimate by default
  m3 with y3@0; !Fix this or Mplus will estimate by default
  x3 with y3@0; !Fix this or Mplus will estimate by default
  x2 on x1*0.71;
  x3 on x2*0.71;
  x2*0.4959;
  x3*0.4959;
  m1*1;
  m2 on m1*0.36;
  m2 on x1*0.25 (a);
  m3 on m2*0.36 x2*0.25;
  M2*0.7650;
  M3*0.7650;
  y1*1;
  y2 on y1*0.51 m1*0.25 x1*0;
  y3 on y2*0.51;
  y3 on m2*0.25 (b);
  y3 on x1*0;
  y2*0.6420;
  y3*0.6420;
MODEL CONSTRAINT:
  NEW(med*0.0625);
  med=a*b;
OUTPUT: TECH9;
SAVEDATA: RESULTS=cell_364.txt;
```


!Mplus code that reads in the data created from the first syntax and
!fits the composite variable model

```

DATA:    FILE=cell_76replist.dat;
        TYPE = MONTECARLO;
VARIABLE: NAMES = mA1 mB1 mC1 mA2 mB2 mC2 mA3 mB3 mC3 x2 x3
          y2 y3 x1 y1;
        USEVARIABLES = x2 x3 y2 y3 x1 y1 m1_c m2_c m3_c;
DEFINE:   m1_c=(mA1+mB1+mC1)/3;
          m2_c=(mA2+mB2+mC2)/3;
          m3_c=(mA3+mB3+mC3)/3;
MODEL:
    x1*1;
    x1 with m1_c*0.2384;
    x1 with y1*0.0663;
    m1_c with y1*0.1389;
    x3 with m3_c@0; !Fix this or Mplus will estimate by default
    m3_c with y3@0; !Fix this or Mplus will estimate by default
    x3 with y3@0; !Fix this or Mplus will estimate by default
    x2 on x1*0.71;
    x3 on x2*0.71;
    x2*0.4959;
    x3*0.4959;
    m1_c*1;
    m2_c on m1_c*0.36;
    m2_c on x1*0.25 (a);
    m3_c on m2_c*0.36 x2*0.25;
    m2_c*0.7650;
    m3_c*0.7650;
    y1*1;
    y2 on y1*0.51 m1_c*0.25 x1*0;
    y3 on y2*0.51;
    y3 on m2_c*0.25 (b);
    y3 on x1*0;
    y2*0.642;
    y3*0.642;
MODEL CONSTRAINT:
    NEW(med*0.0625);
    med=a*b;
OUTPUT: TECH9;
SAVEDATA: RESULTS=cell_652.txt;

```

A SIMULATION STUDY OF TWO-STAGE PIECEWISE LATENT GROWTH
CURVE MODELS FOR TESTING LONGITUDINAL MEDIATION

by

BENTLEY, J. P. AND ROTH, D. L.

In preparation for *Structural Equation Modeling*

Format adapted for dissertation

Abstract

Latent growth curve (LGC) modeling can be used to assess longitudinal mediation, with one such approach being the parallel process model of mediation. Although this approach has several advantages over cross-sectional tests of mediation, a common criticism is that the model cannot be used to establish that prior changes in the mediator are related to future changes in the outcome (i.e., a lack of temporal precedence). Although definitive cause and effect statements still cannot be made about the mediator-outcome relationship, the two-stage piecewise parallel process model of mediation can be used to test whether early growth in a mediator is related to later growth in an outcome (i.e., temporal sequentiality can be established). This article provides an overview of the two-stage piecewise parallel process model of mediation and reports on a simulation study designed to examine the statistical performance of methods used to test mediation in such a model and also to examine the impact of misspecifying a true piecewise model as a single-stage parallel process model of mediation that assumes linear growth trajectories under a variety of conditions. Results demonstrate that fairly large samples, in some cases 1,000 or more, were generally required to minimize bias of mediated effect estimates and to achieve adequate statistical power. Furthermore, under the conditions of the present study, LGC models of mediation are quite sensitive to model misspecifications that fail to account for the true state of temporal precedence, both in terms of model misfit and parameter estimate bias, suggesting that caution should be exercised in the interpretation of single-stage parallel process mediation models without strong theory linking growth in the mediator and growth in the outcome or without prior established evidence of the temporal relationship between mediator and outcome.

Introduction

Latent growth curve (LGC) modeling is an increasingly used technique for the analysis of longitudinal data, allowing researchers to explicitly model growth (or change) over time for individuals (Duncan & Duncan, 2004; Singer & Willett, 2003). This approach is especially useful when one is interested in studying individual differences in change as well as correlates, predictors, and outcomes associated with such change (Raykov, 1998). Another substantial body of literature has explored the concept of mediation, where researchers make an explicit attempt to understand the causal mechanism or the process by which an intervening variable, called a mediator (M), transmits an effect of an independent variable (X) to a dependent variable (Y) (MacKinnon, 2008a). Because of its utility in assessing change over time and the call for the development of longitudinal mediation models due to limitations associated with assessing mediation with cross-sectional data (Maxwell & Cole, 2007), LGC modeling has been expanded to include mediation processes. Perhaps the most straightforward approach is to model growth factors based on repeated measures of X, M, and Y. One method for assessing mediation in such models is by testing whether the growth trajectory (i.e., the slope factor) of X influences the growth trajectory of Y via the growth trajectory of M. Such models are referred to as parallel process models of mediation (MacKinnon, 2008b).

A related LGC parallel process model of mediation was developed by Cheong, MacKinnon, and Khoo (2003) for the situation where X is a time-invariant variable, such as group membership, rather than a time-varying variable measured at multiple waves, like M and Y . An example of this LGC model is provided in Figure 1. Evidence of a causal interpretation of a group membership variable's effect on the growth in the mediator and growth in the outcome is significantly enhanced if the group membership variable (X) represents random assignment to treatment groups. However, causal statements concerning the relationship between the growth trajectories of the mediator and the outcome are limited in the parallel process model of mediation because of the lack of information concerning temporal precedence. In other words, if M and Y are assessed at the same measurement occasions, it is not possible to establish that the growth in M precedes the growth in Y . Although the presence of a temporal sequence alone does not unequivocally establish causality, such a sequence is considered stronger evidence for a causal relationship than a simultaneous relationship (Hill, 1965; Kazdin & Nock, 2003). In the parallel process model of mediation outlined in Figure 1, the relationship between the change in M and the change in Y is correlational in nature (MacKinnon, 2008b; Cheong, 2011); it does not allow one to state that prior changes in M are related to future changes in Y . To make causal inferences about this relationship, and thus subsequently the mediated effect, some have argued for the necessity of a priori strong conceptual theory (Cheong et al., 2003).

Several authors have described the possibility of using a two-stage piecewise parallel process LGC model of mediation to overcome the lack of temporal precedence criticism of the previously described parallel process model of mediation (also called a

single-stage parallel process model) (Cheong et al., 2003; Laurenceau, Hayes, & Feldman, 2007; McKinnon, 2008b; Selig & Preacher, 2009; Cerin, 2010; von Soest & Hagtvet, 2011). In this model, the growth in the mediator and the growth in the outcome can be modeled as occurring in separate phases, allowing for an evaluation of the effect of earlier growth in the mediator on later growth in the outcome. Although levels of the mediator are still not randomly assigned in this model, two-stage piecewise parallel process models of mediation provide significant advantages over single-stage models in that temporal sequentiality can be established, providing a more convincing test of a mediational hypothesis.

Although standard (i.e., single-stage) LGC models can be used to model various trajectory shapes (including growth plateaus and limited early growth followed by late linear growth) by setting different loadings on a growth rate factor (or slope), piecewise LGC models can be specified to explicitly model trajectory shapes that vary over time stage (Flora, 2008). Piecewise LGC models also allow for an evaluation of whether the predictors and outcomes associated with change in a given variable are different at different time points. This suggests that piecewise parallel process models of mediation may provide better estimates of the mediated effect than single-stage models in situations where trajectory shapes for the mediator and the outcome are time dependent (MacKinnon, 2008b). For example, consider a true population model where a prevention program's (i.e., the independent variable) influence on cognitive or attitudinal variables (i.e., the mediators) occurs fairly early followed by a plateau, whereas the ultimate change in behavior (i.e., the outcome) does not occur until some later time during program evaluation.

Due to its relative parsimony, a commonly utilized growth pattern when conducting LGC modeling is the linear trend (i.e., the growth trajectory over time is best represented by a straight line). Although Cheong et al. (2003) recommend exploring the growth patterns in M and Y, and utilizing the most appropriate trajectory when conducting LGC mediation modeling, this recommendation may be ignored. A model misspecification that assumes linear growth trajectories for M and Y when the true population model is a two-stage piecewise model of mediation has the potential to produce significantly biased estimates of the mediated effect.

Several empirical studies have tested longitudinal mediation hypotheses using parallel process models (Cheong et al., 2003; Liu et al., 2009; Audrain-McGovern, Rodriguez, & Kassel, 2009; Roesch et al., 2009; Roesch, Norman, Villodas, Sallis, & Patrick, 2010; Littlefield, Sher, & Wood, 2010) and others have used Monte Carlo simulation methods to assess estimation accuracy and power in the single-stage parallel process model of mediation (Cheong, 2011). However, applied mediation analyses using piecewise parallel process growth models are rare (von Soest & Hagtvet, 2011) and, to our knowledge, no Monte Carlo studies have evaluated the performance of this statistical method for evaluating mediation relationships. One application can be found in the work of Flora, Khoo, & Chassin (2007), who tested a longitudinal mediation hypothesis using data from a cohort-sequential design. However, even though the ultimate dependent variable in their mediation framework was later growth in the outcome (i.e., a second stage heavy drinking growth trajectory), these authors used an intercept factor (i.e., externalizing status at a certain age) rather than a growth rate factor as the putative mediator.

The purpose of this study was to evaluate the statistical performance (bias, power, and Type I error rate) of methods used to test mediation in a two-stage piecewise parallel process latent growth curve model and to examine the impact of misspecifying the true piecewise model as a single-stage parallel process model of mediation under different conditions (i.e., degree of later growth in the mediator, degree of earlier growth in the outcome, complete or partial mediation, size of the mediated effect, and sample size). Prior to a description of the simulation study design and results, LGC modeling (including the specification of piecewise growth models) and the use of LGC models to assess longitudinal mediation are briefly reviewed.

Background

Latent Growth Curve Modeling

In a latent growth curve analysis, observations of the same variable from multiple time points are used to model change over time for individuals. Two parameters are considered latent variables: the intercept (or status) factor, which is commonly used to measure initial or baseline status for each individual (although it can be used to indicate status at any measurement point), and a slope (or growth rate) factor, which indicates individual change over time. The coding of time in LGC modeling is accomplished through specific constrained elements in a factor loading matrix. In matrix form, the unconditional growth model is represented as:

$$\mathbf{Y}_i = \mathbf{\Lambda}\boldsymbol{\eta}_i + \boldsymbol{\varepsilon}_i \quad (1)$$

where \mathbf{Y}_i is a $T \times 1$ vector of repeated measures of the variable Y for individual i over the T time points ($t = 0, 1, 2, \dots, T$), $\mathbf{\Lambda}$ is a $T \times 2$ matrix of factor loadings on the growth

factors, $\boldsymbol{\eta}_i$ is a 2 x 1 vector of latent factors representing the two growth parameters, the intercept and the slope, and $\boldsymbol{\varepsilon}_i$ is a $T \times 1$ vector of measurement errors or residuals.

Values for the loadings for the status factor η_{1i} are constrained to be 1 and values for the loadings of the growth rate factor η_{2i} are fixed by the analyst to reflect time intervals between measurements and the shape of the growth trajectory (e.g., linear). For example, the loadings of [0 1 2 3 4] for η_{2i} reflect a linear trajectory across five time points at evenly spaced intervals (hence that factor is typically labeled “linear slope”). Nonlinear functions, including initial growth followed by a plateau, can be modeled by fixing the factor loadings to certain values (e.g., [0 1 1 1 1]) or allowing some loadings to be freely estimated to reflect the trajectory from the observed data (see Littlefield et al. (2010) for an example).

The means of these latent variables reflect the population average intercept and growth rate, while their variances reflect individual differences. When the variances are significant, the unconditional model can be expanded to include predictors of the two growth parameters. This conditional model in matrix form is represented as:

$$\boldsymbol{\eta}_i = \boldsymbol{\mu}_\eta + \boldsymbol{\Gamma}\mathbf{X}_i + \boldsymbol{\zeta}_i \quad (2)$$

where $\boldsymbol{\mu}_\eta$ is a 2 x 1 vector of regression intercepts, $\boldsymbol{\Gamma}$ is a 2 x K matrix of regression coefficients, \mathbf{X}_i is a $K \times 1$ vector of K individual covariates, and $\boldsymbol{\zeta}_i$ is a 2 x 1 vector of residual (or error) terms.

The Parallel Process Model of Mediation

Growth processes for other variables (i.e., parallel processes) can be added to the basic LGC model, allowing researchers to address many different research hypotheses,

including longitudinal mediation. In such models, it is possible for growth parameters from one process to be regressed on growth parameters from other processes. Figure 1 provides one example of a parallel process model of mediation with the independent variable of interest being a time-invariant variable, such as assignment to treatment or control group, and the mediated effect of interest defined as the effect of X on the growth rate of Y via the growth rate of M ($X \rightarrow S_M \rightarrow S_Y$).¹

Cheong et al. (2003) and MacKinnon (2008b) outline several steps for evaluating parallel process mediation models. This includes evaluating both M and Y for differential growth trajectories, assessing whether the variances of the growth factors are statistically significant, using multiple-groups structural equation modeling (SEM) to evaluate differences in initial status and growth across the levels of the X (e.g., see Kline, 2011), combining the latent growth curve models for M and Y into one parallel process model and then using MIMIC (multiple indicators, multiple causes) modeling (e.g., see Brown, 2006) to assess the impact of the treatment (X) on the growth factors, and finally adding relationships among the growth factors to assess for mediation.

The α coefficient in Figure 1 represents the mean difference in average growth rates between a treatment and control group on the mediator variable (controlling for the effect of the initial status of the outcome), whereas the β coefficient signifies the relationship between the growth rate of the mediator and the growth rate of the outcome, adjusted for the effect of the treatment (and the initial status of the mediator). The τ' coefficient indicates the direct effect of X on the growth in the outcome, or the effect of X on the growth in the outcome that is not correlated with the growth rate of the

¹ Note that I_M , S_M , I_Y , and S_Y are used to represent the growth factors in Figure 1 rather than η_1 , η_2 , η_3 , and η_4 .

mediator. The point estimate of the longitudinal mediated effect of interest is $\hat{\alpha} * \hat{\beta}$ (Cheong et al., 2003). Although there are others (MacKinnon, Lockwood, Hoffman, West, & Sheets, 2002), two commonly used methods to test for statistically significant mediation are the joint significance test (Cohen & Cohen, 1983) and the Sobel test (1982). When the effects being multiplied together to produce the mediated effect are both statistically significant, the joint significance test concludes that there is a significant mediated effect. In the parallel process model of mediation in Figure 1, this would include tests of the α and β paths. No estimate of the mediated effect is necessary and such a test does not provide confidence intervals.

The Sobel standard error formula for the $\hat{\alpha} * \hat{\beta}$ mediated effect is:

$$\sqrt{\hat{\alpha}^2 s_{\hat{\beta}}^2 + \hat{\beta}^2 s_{\hat{\alpha}}^2} \quad (3)$$

where $s_{\hat{\alpha}}$ and $s_{\hat{\beta}}$ are the standard errors for the α and β estimates. The Sobel standard error is implemented in several statistical software programs and can also be used to construct confidence intervals for the mediated effect, in addition to significance testing. As others have noted, the distribution of a product of two independent normally distributed random variables is generally not normally distributed (MacKinnon, 2008a). Thus, the use of the Sobel standard error relies on asymptotic theory (i.e., normal-theory confidence limits and hypothesis tests), and the net result is generally conservative hypothesis tests (low Type I error rates and low power) and confidence intervals (i.e., empirical coverage probabilities larger than 95% for a 95% confidence interval).

Although the putative mediator in the parallel process model depicted in Figure 1 is the slope, it is possible to test mediated effects involving the mediator's intercept factor

(von Soest & Hagtvet, 2011). The intercept or status factor is the model-implied value of the repeatedly measured variable when the loading on the growth rate factor is 0. While often times this is the value of the variable at baseline, this can be changed so that this factor reflects status at any time point, not just initial status, by selecting a new reference point. Thus, the loadings of $[0 \ 1 \ 2 \ 3 \ 4]$ for η_{2i} can be changed to $[-1 \ 0 \ 1 \ 2 \ 3]$, which will not change the interpretation of the growth rate factor, but now the status factor is the value of the variable at $t = 1$ rather than baseline. As noted by Selig & Preacher (2009), caution should be exercised so that mediation paths that run contrary to time are not tested. Thus, using M status at the last time point as a mediator for the growth in Y occurring over all time points should not be tested as some growth in Y would occur before the final M status is achieved.

Piecewise Latent Growth Curve Modeling

LGC models can be extended to include additional growth factors. Perhaps the simplest of these models is the inclusion of a second slope (or growth rate) factor to reflect change over a second time segment, the two-piece growth model (Flora, 2008). Using the matrix form of the unconditional growth model in equation 1, η_i is now a 3×1 vector of latent factors representing the three growth parameters (the intercept, early growth rate, and later growth rate) and Λ is a $T \times 3$ matrix of factor loadings on the three growth factors. An example of such a factor loading matrix with five measurement occasions is:

$$\Lambda = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 2 & 0 \\ 1 & 2 & 1 \\ 1 & 2 & 2 \end{bmatrix} \quad (4)$$

where the intercept factor represents status at baseline or the first measurement occasion, the first growth rate factor is the linear change over the first three time points, and the second growth factor represents linear change over the last three time points. The conditional model is a straightforward extension, where predictors of all three growth parameters can be included (i.e., the Γ matrix in equation 2 becomes a $3 \times K$ matrix of regression coefficients).

The transition point, or the point at which the two pieces are joined, is referred to as the “knot.” In some applications of piecewise models, the knot represents a point where the slopes representing growth on the variable begin to change and thus piecewise models can be used to model nonlinear trajectories. In other applications, the knot may represent other considerations, such a transition from active treatment to a maintenance period or from middle school to high school (Sayer & Willett, 1998; Khoo, 2001; Duncan & Duncan, 2004; Flora, 2008). In these applications, it is possible that the growth over the entire time period may be linear, but the goal may be to assess different predictors of growth during the two time periods.

The Two-Stage Piecewise Parallel Process Model of Mediation

As with the LGC model, growth processes for other variables can be added to the basic two-stage piecewise LGC model, potentially allowing for a more convincing test of mediation by allowing an assessment of the effect of earlier growth in the mediator on

later growth in the outcome. Such a model is presented in Figure 2 where the primary mediated effect is $\alpha*\beta$, and is defined as the effect of X on the growth rate of Y occurring at a later stage via the growth rate of M occurring at an earlier stage ($X \rightarrow S1_M \rightarrow S2_Y$). Of note is that there are several other possible mediated effects. For example, one may elect to examine the relationships between earlier growth in M and earlier growth in Y (mediated effect = $\alpha*\gamma_9$) or later growth in M and later growth in Y (mediated effect = $\gamma_3*\gamma_{10}$) (MacKinnon, 2008b). Although these “contemporaneous” mediated relationships still suffer from a lack of temporal precedence, they nevertheless may be of substantive interest.

Other mediated effects may involve the intercept (or status) factors. Given that it is possible to change the loadings in Λ such that the intercept factor reflects status at any time point, not just initial status (see previous discussion), it is possible to change the nature of a mediated effect involving the intercept by changing the time used to define the intercept (Selig & Preacher, 2009). One can allow temporal precedence in such models with the intercept as the putative mediator. For example, one might test whether the effect of X on later growth in Y (i.e., times 3, 4, and 5) is mediated via status of M at time 2. This was the approach taken by Flora, Khoo, & Chassin (2007), primarily because the variance of the growth rate factor for their proposed mediator was not statistically significantly different from zero.

In summary, piecewise growth models provide researchers with the flexibility to examine nonlinear growth trajectories, antecedents and consequents of change at different growth periods, and the ability to test a number of different mediational

hypotheses, some of which specifically address the criticism of the lack of temporal precedence associated with the single-stage parallel process model of mediation.

Methods

Population Model

The present simulation study sought to evaluate the performance of estimating and testing mediation in a two-stage piecewise parallel process LGC model and to evaluate the effects of misspecifying the true piecewise model as a single-stage parallel process LGC model. Thus, the population longitudinal mediation model was based on the model in Figure 2, a model with five measurement occasions for M and Y (both simulated to be continuous), and an independent variable, X, designed as a dichotomous, time-invariant variable with a 50/50 split, to mimic a randomized controlled trial with a treatment and a control group with equal treatment allocation. The factor loading matrix, Λ , for the population model was that specified in equation 4 for both the mediator and the outcome, reflecting equally-spaced intervals between measurement occasions. Thus, I_M and I_Y represent initial status (i.e., at baseline) for the mediator and the outcome, respectively; $S1_M$ and $S1_Y$ represent linear growth rates for the mediator and outcome, respectively, at an earlier stage of the study; and $S2_M$ and $S2_Y$, represent linear growth rates for the mediator and outcome, respectively, at a later stage of the study.

Slope was considered the putative mediator, thus the mediated effect in this study was defined as the independent variable (X) affecting the growth rate of Y occurring at a later stage via the growth rate of M occurring at an earlier stage ($X \rightarrow S1_M \rightarrow S2_Y$) and was estimated as the product of these two paths in subsequent analyses ($\hat{\alpha} * \hat{\beta}$). No

contemporaneous mediational relationships (i.e., $X \rightarrow S1_M \rightarrow S1_Y$) were specified, as γ_9 and γ_{10} were fixed to zero. In addition, γ_1 and γ_2 (i.e., the relationship of X to initial status of the mediator and outcome) were both set to 0 to mimic a randomized trial.

Other parameters were selected based on prior simulation studies involving LGC models (Muthén & Curran, 1997; Muthén & Muthén, 2002; Hertzog, von Oertzen, Ghisletta, & Lindenberger, 2008; Thoemmes, McKinnon, & Reiser, 2010; Cheong, 2011). The means of the initial status factors for both the mediator and outcome were set to 0 with variances of 1. The residual variances of observed M_1 through M_5 and Y_1 through Y_5 were set to be same (0.25), a simplifying assumption that has been used in several LGC modeling simulation studies (Muthén & Muthén, 2002; Hertzog et al., 2008; Thoemmes et al., 2010). This yields a model-implied total variance for M_1 and Y_1 of 1.25 and an average reliability (R^2) of these variables (i.e., at baseline) of 0.8. The R^2 for the measures at the other time points vary by the other parameters in the model, but are all fairly close to 0.8.

Covariances among the initial status factors and disturbances of the first growth factors for both the mediator and the outcome (i.e., $\text{Cov}(I_M, S1_M)$ and $\text{Cov}(I_Y, S1_Y)$) were set at -0.1. Because there is more time separating the initial status factors and the disturbances of later stage growth rate factors, these covariances (i.e., $\text{Cov}(I_M, S2_M)$ and $\text{Cov}(I_Y, S2_Y)$) were set at -0.05. For reasons of clarity, none of these covariances are shown in Figure 2. Relationships between initial status factors of one process and the growth rate factors of the other process (i.e., γ_5 , γ_6 , γ_7 , and γ_8 in Figure 2) were all set at -0.1. All of these negative relationships are based on the assumption often observed in applied LGC analyses that “individuals with higher levels at the initial measurement

point showed smaller changes at the later points” (Cheong, 2011, p. 201). Covariances of the disturbances of the growth rate factors (i.e., $\text{Cov}(S1_M, S2_M)$ and $\text{Cov}(S1_Y, S2_Y)$) were set at -0.05, smaller than the covariance involving initial status and the first growth rate factor based on previous empirical studies using piecewise LGC modeling (Lee & Rojewski, 2009). The covariance among the initial status factors for the mediator and outcome was fixed at 0.3 (Cheong, 2011).

For both the mediator and the outcome, the average growth for the first stage (i.e., $S1_M$ and $S1_Y$) and the second stage (i.e., $S2_M$ and $S2_Y$) were set to 0, with residual variances of 0.1 for these growth rate factors. This simplifying assumption indicates no growth over time on average for the control group (i.e., model-implied means of 0 at all five measurement occasions) for both the mediator and the outcome. For the treatment group, growth is a function of the other parameters in the model and the model-implied means for the treatment group are fully described below in the section describing simulation study conditions.

Path α in Figure 2 expresses early growth for the treatment group above the control group on the mediator and was fixed in all study conditions at 0.4. A single value for this path was selected for this study because this effect is often well supported in the literature prior to an empirical study and the experimental manipulation (i.e., the X variable in Figure 2) is often designed to have a strong effect on the mediator. Thus, investigators designing longitudinal studies modeled after Figures 1 or 2 often have strong evidence for a treatment effect on the mediator of interest before the study is conducted. Given the other parameters already fixed in the population model (i.e., the residual variance of $S1_M$, γ_6 , the variance of I_Y , and the variance of X (simulated to be

50/50 dichotomy, therefore the variance of X is 0.25)), a value of 0.4 for α corresponds to a unique explained variance of path α (for the prediction of $S1_M$) of just over 26%, a large effect (Cohen, 1988). In addition, the ratio of the difference in slope means for the two values of X divided by the standard deviation of $S1_M$ is about 1, which is also indicative of a large effect (Cohen, 1988).

Simulation Study Conditions

Mplus 6.11 (Muthén & Muthén, Los Angeles, CA) was used for conducting simulations and subsequent model fitting. Source code for data generation for one of the combinations of the experimental conditions along with code for the fitting of the two model specifications can be found in the Appendix. A total of six factors were manipulated in this simulation study: 1) two model specifications, 2) four sizes of the mediated effect (which is based on path β since path α was fixed in all simulation conditions), 3) two levels of the size of path τ' (to reflect complete and partial mediation), 4) two levels of later growth in the mediator, 5) two levels of earlier growth in the outcome, and 6) five different sample sizes. Of these, only five contributed to the creation of different sets of replications (model specification can be considered a within-subjects factor, where both levels of this factor were applied to all replications in each condition). Thus, data were generated under 160 unique experimental conditions ($4 \times 2 \times 2 \times 2 \times 2 \times 5$).

A total of 1,000 replications were simulated for each of these 160 conditions and the properly specified model (i.e., the model outlined in Figure 2) was fit to the 1,000 raw datasets in each condition using maximum likelihood estimation. Out of these 1,000

replications in each of the 160 conditions, replications that failed to converge or resulted in negative variances or residual variances were excluded, and then the first 500 replications with converged and proper solutions were selected for subsequent analyses (Paxton, Curran, Bollen, Kirby, & Chen, 2001), providing a total of 80,000 data sets. In a few experimental conditions, there were close to 400 replications out of the 1,000 with improper solutions or a failure to converge. Almost all of the problematic replications occurred with $n = 100$ or $n = 200$ (mostly, $n = 100$). Thus, only 500 converged and proper solutions from the perfectly specified model were used and considered for the misspecified model. Thus, the two model specifications described below were fit to the 80,000 simulated data sets. Some of these replications, although resulting in converged and proper solutions in the perfectly specified model, failed to converge or resulted in improper solutions in the misspecified model. Such problems were encountered in 2,078 replications (about 2.6% of the replications), including 861 cases of failed convergence and 1,217 cases of out-of-range values. Although some of these occurred when $n = 200$, almost all occurred with $n = 100$. These replications were excluded from subsequent analyses, leading to a slightly unbalanced design.

Model specification. Two different models were specified and were fit to each data set. The first model specification was the properly specified two-stage piecewise parallel process model of mediation (Figure 2) (i.e., the estimated model perfectly corresponded to the population model). The second model specification was a single-stage parallel process model of mediation. In this model, only one growth rate factor was estimated for both the mediator and the outcome and both growth trajectories were modeled as linear (Figure 1). The slope was still considered the putative mediator, thus

the mediated effect in the misspecified model was defined as the independent variable (X) affecting the growth rate of Y via the growth rate of M ($X \rightarrow S_M \rightarrow S_Y$). Maximum likelihood was used as the estimation method for both specifications.

Size of the mediated effect. Because path α was fixed at 0.4 under all simulation conditions, path β was varied to manipulate the size of the mediated effect. The values of 0, 0.15, 0.40, and 0.65 were chosen for the path β to reflect zero, small, medium, and large effects for the mediated effect. Given the other parameters already fixed in the population model (i.e., the residual variance of S_{2Y} , γ_7 , the variances of S_{1M} and I_M , and a number of covariances), these values correspond to unique explained variance of the β path of 0%, 2.8%, 16.2%, and 32.2% (when $\tau' = 0$) and 0%, 2.4%, 13.0%, and 25.9% (when $\tau' = 0.25$), which correspond to 0, small, moderate, and large effects on the R^2 metric (Cohen, 1988).

Size of path τ' . Two levels for path τ' (or the direct effect of X) were used to reflect complete ($\tau' = 0$) and partial ($\tau' = 0.25$) mediation. The value of $\tau' = 0.25$ corresponds to a small to medium effect in terms of unique explained variance.

Later growth in the mediator. Path γ_3 (Figure 2) was used to manipulate the degree of later growth in the mediator. Given path α is set at 0.4, when $\gamma_3 = 0$, the growth in mediator due to treatment is early (and linear) with no additional growth beyond time 3 (i.e., a stabilized plateau is reached) and when $\gamma_3 = 0.4$, there is later growth in the mediator due to treatment and the total growth is basically the same over the five periods (it is essentially linear).

Earlier growth in the outcome. Path γ_4 (Figure 2) was used to manipulate the degree of earlier growth in the outcome. Given a treatment effect on S_{2Y} (i.e., $\beta \neq 0$

and/or $\tau' \neq 0$), when $\gamma_4 = 0$, the growth in outcome due to treatment is late with no growth in the first three time points and when $\gamma_4 = 0.4$, the growth in the outcome due to treatment occurs early (and is linear over the first three measurement occasions) and there is late growth as well from the contributions of β and τ' . In the latter case, the total amount of change over time is a function of β and τ' .

Thus, the total growth over the entire study period in the mediator is function of α and γ_3 , while the total growth in outcome is function of α , β , τ' , and γ_4 . Table 1 provides model-implied means for the measured variables M_1 - M_5 and Y_1 - Y_5 under a few different experimental conditions.

Sample size. Five sample sizes often encountered in social and behavioral science research were considered: 100, 200, 500, 1000, and 2000.

Study Outcomes

Fit indices. Two commonly used indices were used to assess the fit of the two model specifications to the observed data for each replication: the root mean square error of approximation (RMSEA) and the comparative fit index (CFI). The means for these two indices across all replications for each experimental condition were calculated and compared to literature-based cut-off criteria (Bandalos, 2006). Good model fit is indicated by an RMSEA of 0.05 or less (Brown & Cudeck, 1993) and a CFI above 0.95 (Hu & Bentler, 1999).

Parameter and standard error estimation. The accuracy of point estimation in the two-stage piecewise parallel process model of mediation, as well as the effects of model misspecification on estimation accuracy, were evaluated by examining both bias

and relative bias. While the primary interest was the estimate of the mediated effect, other parameters in the model were examined as bias associated with estimation of these parameters may help explain the effects of model misspecification on the bias in the mediated effect. Relative bias is defined as the ratio of the deviation in an estimate from the true value to the true value:

$$\frac{\hat{\theta} - \theta}{\theta} \quad (5)$$

$\hat{\theta}$ is the point estimate of the effect of interest from the simulated data whereas θ refers to the true value specified in the population model. Absolute values greater than 0.10 were considered problematic (Kaplan, 1988). Under conditions where the true value was zero, the bias (the numerator in equation 5) was calculated and reported.

Several studies report on the relative bias associated with the estimation of the Sobel (1982) standard error (e.g., see MacKinnon et al., 2002; Taylor, MacKinnon, & Tein, 2008; Cheong, 2011) due to questions concerning its performance especially with smaller samples. To calculate this relative bias, equation 3 was used to calculate the Sobel (1982) standard error estimate for each of the 500 replications under a given set of conditions. The standard deviation of the mediated effect across the same 500 replications was used as the true value of the standard error and then equation 5 was used to calculate relative bias.

Power and Type I error rates. Effects of interest were tested for significance in each of the 500 replications under all experimental conditions using a two-tailed test at the 0.05 level of significance. Both the Sobel test and the joint significance test were used for testing the mediated effect. Under conditions when the true effect was 0, the proportion of replications in which the null hypothesis was rejected was used as an

estimate of the Type I error rate. When the true effect was nonzero, this proportion was used as the measure of power.

Analysis

SAS 9.2 (SAS Institute, Cary, NC, USA) was used for analyzing the simulation results. To examine the main effects and interactions of the six design factors on the study outcomes, analysis of variance was used for continuous outcomes (bias, relative bias, fit indices) and logistic regression was used for dichotomous outcomes (Type I error and power). Because the number of observations was so large, effects were interpreted on the basis of effect sizes rather than conventional $p < 0.05$ significance testing (Paxton et al., 2001). The proportion of total variation accounted for (η^2) was used as the measure of effect for ANOVA models and the proportional reduction in deviance attributable to a predictor (R_L^2) was used for logistic regression models. To facilitate estimation in the logistic regression models, size of the mediated effect and sample size were treated as centered, quantitative predictors (see Taylor et al., 2008 for a similar approach). The two model specifications technically comprise a within-subjects factor; however, model specification was treated as a between-subjects factor in the analysis because the results were interpreted in terms of effect sizes rather than significance testing. Given the slight differences in cell sizes due to lack of convergence or improper solutions for the misspecified model for conditions where sample sizes were small (i.e., there was an unbalanced design), least-squares means (i.e., group-based averages or unweighted means) were used when collapsing across cells in the reporting of results (Keppel & Wickens, 2004).

Results

Model Fit

Overall, the study factors (and all possible interactions) accounted for 88.5% and 89.7% of the total variation in RMSEA and CFI, respectively. For RMSEA, most of the explained variation (98%) was due to the main effects of, and interactions among, four factors: model specification, size of the mediated effect, later growth in the mediator, and sample size, while for CFI, the first three of these factors and their interactions (i.e., not sample size) accounted for most of the explained variation (93%). Model specification had the single largest effect for both RMSEA and CFI, as expected. Mean fit index values are presented in Table 2 as a function of these four factors, collapsing across levels of the size of the direct effect and degree of earlier growth in the outcome. For the properly specified model, mean RMSEA and CFI values indicated good fit under all conditions, with a slight dependency on sample size for RMSEA (larger sample sizes are associated with decreasing values of RMSEA; this was not unexpected, see Kline, 2011). The misspecified model was generally associated with significantly poorer fit. RMSEA was generally sensitive to detecting misfit associated with model misspecification under all conditions (range: 0.068 to 0.105), with higher values (i.e., worse fit) generally associated with larger mediated effects and conditions where there is no late growth in the mediator (i.e., a stabilized plateau). Alternatively, mean values for CFI indicated acceptable fit under several conditions (range: 0.927 to 0.972), although the general pattern observed for RMSEA held; that is, worse fit is generally associated with larger mediated effects and conditions where there is no late growth in the mediator.

Parameter Estimation

The study factors (and all possible interactions) collectively accounted for 12.8% of the total variation in the relative bias associated with the mediated effect estimates, with most of the explained variation (99%) attributed to four factors and their interactions: model specification, size of the mediated effect, later growth in the mediator, and sample size. Mean relative bias values for the mediated effect are presented in Table 3 as a function of these four factors, collapsing across collapsing across levels of the size of the direct effect and degree of earlier growth in the outcome. For the properly specified model, in cases of small effects, the relative bias was generally less than 0.10 only when the sample size was 1,000 or larger. As the size of the mediated effect increased, the relative bias tended to decrease such that the relative bias for the mediated effect in the properly specified model was less than 0.10 across all sample sizes for the largest effect. Using a single-stage parallel process model of mediation with growth trajectories for both the mediator and the outcome modeled as linear led to significant bias in the estimation of the mediated effect in just about all conditions. In some conditions the effect was underestimated and in other conditions it was overestimated. Even in situations where the true total growth for the mediator for the treatment group over the five periods is essentially linear (i.e., $\gamma_3 = 0.4$), there was still a significant degree of bias in the estimate of the mediated effect.

An examination of bias associated with the estimation of the mediated effect when the true mediated effect was zero (i.e., when relative bias could not be calculated) revealed similar findings as the analysis of the relative bias when the true mediated effect was nonzero. Model specification, later growth in the mediator, and sample size

accounted for most of the explained variation (98%, with total variation accounted by all study factors of 12.6%). Significant overestimation of the zero mediated effect occurred in the misspecified models (Table 4).

Only two study factors accounted for significant variation in the relative bias of the path α estimates: model specification and later growth in mediator (i.e., size of path γ_3). These two factors and their interaction accounted for about 99% of the explained variation (the study factors and all possible interactions collectively accounted for 78.5% of the total variation). For the properly specified model, the relative bias was near zero for both levels of path γ_3 . However, for the misspecified model, path α is underestimated on average by about 50% when path γ_3 is 0 (i.e., no late growth for the mediator in the true population model), reflecting a forced linear fit to a stabilized plateau. When path γ_3 is 0.4 in the true population model (i.e., when there is late growth as well as early growth in the mediator), the relative bias for the path α estimates is essentially 0 (-0.005).

The study factors together with all possible interactions accounted for a significant amount of total variation in the relative bias of path β estimates when this path was nonzero in the population and in bias of path β estimates when $\beta = 0$ in the population (15.6% and 14.9%, respectively). As with the relative bias associated with estimates of the mediated effect, most of the explained variation for the relative bias of path β estimates (99%) can be attributed to four factors and their interactions: model specification, size of the mediated effect, later growth in the mediator, and sample size. Model specification, later growth in the mediator, and sample size accounted for most of the explained variation (99%) in the bias for path β estimates when the true value of $\beta = 0$, which is also similar to bias in mediated effect when the true mediated effect was zero.

Tables 5 and 6 present the patterns of means associated with the relative bias and bias of the path β estimates, respectively. The findings for the properly specified model closely paralleled the findings for relative bias associated with the mediated effect estimates in Table 3, which is not surprising given that the relative bias for path α estimates was generally near zero for the properly specified model. Likewise, the patterns of bias for the path β estimates when $\beta = 0$ in the properly specified model paralleled for the findings for the mediated effect when the true mediated effect = 0 (Table 4). For the misspecified model, path β is significantly overestimated in all conditions (including the condition of a zero mediated effect), but the degree of bias generally improved with larger mediated effects and larger sample sizes.

Although the degree of earlier growth in the outcome and the size of the direct effect, τ' , did not account nontrivial variation in parameter estimate bias or relative bias for the mediated effect or paths α and β , these factors did account for meaningful variation in the bias of the direct effect estimates. Although it is possible to calculate relative bias when $\tau' = 0.25$, for consistency, the analysis was performed on the bias associated with this parameter estimate. The study factors taken together accounted for 26.2% of the total variation in the bias of the τ' path estimates, with the main effects of, and interactions among, model specification, size of the direct effect, later growth in the mediator, earlier growth in the outcome, and sample size accounting for most of the explained variation (97%). Tables 7 presents mean bias values for the direct effect estimates as a function of these five factors, collapsing across levels of the size of the mediated effect. The main effect of the size of the direct effect was simply confirmation of that manipulation. For the most part, the properly specified model showed no

significant bias associated with this path. However, the direction and magnitude of the bias of the direct effect estimates for the misspecified model varied according to the degree of later growth in the mediator and the degree of earlier growth in the outcome. A greater degree of earlier growth in the outcome (i.e., $\gamma_4 = 0.4$) in the population model was generally associated with a positive bias for the direct effect. However, a greater degree of later growth in the mediator (i.e., $\gamma_3 = 0.4$) was generally associated with a negative bias for the direct effect. In some cases, these effects appear to “cancel out” each other leading to estimates for the direct effect that appear to be close to the true population value (which also helps to explain the counterintuitive finding in Table 7 that for a few condition combinations, increasing sample size appears to be contributing to increasing bias).

Standard Error Estimation

The relative biases of the Sobel standard error of the mediated effect can be found in Table 8. Because the main effects of, and interactions among, model specification, size of the mediated effect, and sample size accounted for most of the explained variation (88%, with total variation accounted by all study factors of 1.1%), the results in Table 8 are collapsed across levels of the size of the direct effect, degree of later growth in the mediator, and degree of earlier growth in the outcome. Sample size had by far the largest effect, which is evident in Table 8, which shows that the Sobel standard error tended to be positively biased under both model specifications and at all mediated effect sizes, when the sample size was 100. The absolute value of the relative bias of the Sobel standard error estimator was generally below 0.10 with sample sizes greater than 100,

although there were some notable exceptions for the properly specified model when $n = 200$.

Type I Error and Power

The Type I error rates for the mediated effects (i.e., when the true mediated effect was zero) using both the Sobel test and the joint significance are shown in Table 9. The study factors accounted for a 41.5% deviance reduction for the Sobel test and 39.6% for the joint significance test, with most of the contribution to this reduction (99% for both tests) coming from the main effects of model specification and sample size, along with a sizable interaction between these two factors. Table 9 therefore presents Type I error rates different levels of these factors collapsing across the other factors. For the properly specified model, Type I error rates were below the nominal level of 0.05 for all sample sizes regardless of method of testing (although Type I errors rates were generally slightly higher for the joint significance test especially at smaller sample sizes). However, because of the sizable bias associated with estimates of the mediated effect in the misspecified model when the true effect was zero (Table 4), the Type I error rate for the mediated effect significantly exceed nominal levels especially at sample sizes of 500 or larger, approaching 0.80 with a sample size of 2,000.

The results with respect to empirical power for the mediated effect can be found in Table 10. In a model using all study factors and interactions, a deviance reduction of 61.8% was achieved for the Sobel test and 59.1% for the joint significance test, with most of the contribution to this reduction (99% for both tests) coming from the main effects of, and interactions among, model specification, effect size of the mediated effect, and

sample size. Table 10 therefore presents empirical power levels for different levels of these factors. In general, regardless of model specification and method used to test mediation, power generally increased with increasing sample size and increasing effect size of the mediated effect, as would be expected. Under the same conditions, power was generally higher with the joint significance test when compared to the Sobel test. Even with proper model specification, sample sizes of 100 or 200 are generally inadequate (i.e., failed to achieve a power of 0.8) to detect a mediated effect, no matter the size, when using a two-stage piecewise parallel process model of mediation. Furthermore, the power to detect small effects was generally low for both the Sobel test and the joint significance test even at a sample size of 2,000. The apparent power advantage for the misspecified model is directly related to the inaccurate estimates of the mediated effect under conditions of model misspecification (Table 3).

Table 11 provides the Type I error rates (when $\tau' = 0$) when and empirical power (when $\tau' = 0.25$) associated with the test of the direct effect. For both Type I error rate and empirical power, the study factors accounted for a nontrivial deviance reduction, 43.7% and 41.1%, respectively. Most of these effects (about 95% for both Type I error rate and power) were attributed to model specification, later growth in the mediator, earlier growth in the outcome, and sample size (and their interactions). Thus, the entries in Table 11 are mean values collapsed across the size of the mediated effect. Type I error rates for the direct effect generally followed the bias associated with the direct effect presented in Table 7. Thus, inflated Type I error rates often occurred in the misspecified model, but were generally near the nominal level (0.05) for the properly specified model. Given the direction of the bias noted in Table 7, a Type I error may be associated with

incorrectly concluding that there is negative or a positive direct effect, depending on the study conditions. With respect to empirical power, in the properly specified model, later growth in the mediator (i.e., path $\gamma_3 = 0.4$) is generally associated with lower power. Because there is no significant bias associated with this path with proper model specification (Table 7), this effect is explained by the increased variability in the estimate for the direct effect in the presence of late growth in the mediator. For the misspecified model, empirical power levels are related to the degree of bias associated with estimation of this path. For example, when there is no early growth in the outcome but there is late growth in the mediator, the direct effect estimates are biased towards zero (Table 7) and thus the power for detecting this effect is quite low (range: 0.046 to 0.11 depending on sample size).

Discussion

Assessing mediational hypotheses with LGC modeling has certain advantages over cross-sectional tests of mediation. The assessment of change over time in both the mediator and the outcome, as well as the ability to estimate individual differences in change, provide more convincing evidence of mediation than single moment-in-time view of the variables involved in a proposed mediation relationship. However, the single-stage parallel process LGC model of mediation suffers from an inability to claim temporal precedence. In other words, it cannot be used to establish that prior changes in M are related to future changes in Y. Although definitive cause and effect statements still cannot be made about the mediator-outcome relationship, the two-stage piecewise parallel process LGC model of mediation does allow one to establish the casual criterion

of temporal sequentiality, in that one can test whether early growth in M is related to later growth in Y. Other strengths of the piecewise models for testing mediation include the ability to specify nonlinear trajectories, increased flexibility in using the intercept factor as mediator, and the ability to model bidirectional or reciprocal causal relationships (e.g., does early growth in the outcome predict late growth in the mediator?). Although single-stage parallel process models have a great deal of flexibility to model trajectory shapes other than linear forms (e.g., see Cheong et al. 2003; Littlefield, et al., 2010), they do not possess the other properties of piecewise parallel process models.

Several authors have recommended the use of piecewise LGC models to assess mediation relationships, however there is little information concerning this model in either the applied or methodological literature. In addition, for reasons of parsimony, researchers may elect to estimate single growth-rate parameters based on a linear form for both the mediator and the outcome, a model misspecification that may produce significant model misfit biased parameter estimates. Therefore, this study examined the statistical performance of methods used to test mediation in a two-stage piecewise parallel process latent growth curve model and to examine the impact of misspecifying a true piecewise model as a single-stage parallel process model of mediation that assumes linear growth trajectories under a variety of conditions.

With respect to statistical performance, although accuracy and power were strongly related to the size of the mediated effect, it is important to note that large samples, in some cases 1,000 or more, were generally required to minimize bias of mediated effect estimates and to achieve adequate statistical power in the two-stage piecewise parallel process model of mediation. Although not tested in the present study,

it is reasonable to suggest two factors that should enhance statistical power for detecting the mediated effect in the piecewise parallel process of mediation on the basis of other methodological research conducted on LGC models: 1) increasing R^2 of the measured variables (or growth curve reliability – GCR) (Hertzog et al., 2008; Cheong et al., 2011) and 2) increasing the number of measurement occasions (Muthén & Curran, 1997; Cheong, 2011). In the present study, GCR estimated at time 1 was 0.8, a fairly sizable amount. Although using highly reliable measures for the mediator and outcome should enhance GCR, it is also possible to use a measurement model to extract latent variables from multiple fallible indicators and subsequently fit a growth curve to these latent variables (which are theoretically without error). Such models go by a variety of names including curve-of-factors models (McArdle, 1988; Duncan & Duncan, 2004; Liu et al., 2009) and second-order latent growth models (Hancock, Kuo, & Lawrence, 2001; Sayer & Cumsille, 2001). Future studies using such methods in the piecewise parallel process mediation modeling framework are warranted.

Increasing the number of measurements is also likely to improve the power to detect a mediated effect. In the present study, five measurement occasions were used. Indeed, this is the minimum number of measurements necessary for identification of a two-piece latent growth curve model that has no overlap between adjacent linear growth segments (Bollen & Curran, 2006; Flora, 2008). Given limited resources and the logistical difficulties associated with longitudinal studies, securing five measurements may be problematic, much less any additional measurements needed to enhance statistical power. Some have suggested this as a reason for the limited number of applications

using piecewise growth models to test mediation (von Soest & Hagtvet, 2011), despite some clear advantages to their use.

With respect to the second study purpose, under the conditions of the present study, LGC models of mediation are quite sensitive to model misspecifications that fail to account for the true state of temporal precedence, both in terms of model misfit and parameter estimate bias. Model fit indices generally revealed poor fit for the misspecified models, which was worse under conditions of a stabilized plateau for the mediator. This was not unexpected as the misspecified model forced a linear growth rate over all five time points. RMSEA was generally more sensitive to detecting misfit than CFI. Others have observed that the CFI may be more sensitive to misspecified factor loadings (Heene, Hilbert, Draxler, Ziegler, & Bühner, 2011; Kline, 2011), which may explain its failure to reject misfit under some conditions in the present study. There is a growing body of literature calling into question the applications of general cutoff values for fit indices under all conditions (Marsh, Hau, & Wen, 2004; Fan & Sivo, 2005, 2007; Yuan, 2005; Heene et al., 2011).

Using a single-stage parallel process model of mediation with growth trajectories for both the mediator and the outcome modeled as linear when the true model was a true-stage piecewise model led to significant bias for several parameter estimates in just about all conditions. In general, path β , which relates growth in the mediator to growth in the outcome, is overestimated under the model misspecification with single linear growth parameters for M and Y. Contrast that with the bias associated with estimates of path α , which is generally unbiased when there is late growth as well as early growth in the mediator (i.e., the underlying pattern over all five measurement occasions is essentially

linear), and strongly negatively biased when the growth of the mediator reaches a plateau after three measurement occasions. These results lead to some interesting findings with respect to the bias of the mediated effect estimates. The downward bias in path α under certain conditions appears to “counteract” the upward bias associated with path β in terms of the mediated effect estimates; thus, in some conditions the mediated effect was underestimated and in other conditions it was overestimated. This pattern of differential bias associated with estimates of the two paths comprising the mediated effects explains why in some situations under model misspecification the mediated effect estimates appear unbiased and actually become more biased with increasing sample size.

It is important to note that even under conditions where the true total growth for the mediator for the treatment group over the five periods is essentially linear (i.e., $\gamma_3 = 0.4$), there was still a significant degree of bias in the estimate of the mediated effect. This would suggest that even if the nonlinear pattern for growth (i.e., the stabilized plateau) was appropriately modeled in a single-stage parallel process model (e.g., using [0 1 2 2 2] as the loadings for the growth rate parameter), there would still be bias in the mediated effect estimates if one does address temporal precedence. The parameter estimate bias for the mediated effect leads to a predictable pattern of Type I error rates. Thus, given a large enough sample, it is very possible to conclude that a mediated effect is statistically significant when the true mediated effect in the population with temporal precedence is zero, even in cases where total growth over all measurement occasions appears to be linear.

The pattern of bias of the direct effect estimates under the misspecified model, while somewhat dependent on the degree of later growth in the mediator, is also

influenced by the degree of earlier growth in the outcome, albeit these biases are in opposite directions. Thus, depending on the combination of conditions, this bias may be positive, negative, or have the appearance of zero bias. It is interesting to note that in the case of complete mediation, under several conditions the parameter estimate for the direct effect is negative (i.e., negatively biased) but the mediated effect estimate is positive (i.e., positively biased). Depending on the circumstances this may be misinterpreted as a case of inconsistent mediation (MacKinnon, Krull, & Lockwood, 2000).

All of these findings suggest caution should be exercised when interpreting the results of single-stage parallel process models of mediation. Some have argued that strong theory linking the two growth rates (i.e., the mediator and the outcome) in the single-stage parallel process model of mediation is a requirement for causal inference (Cheong et al., 2003), while others have recommended that such models only be used when “a temporal relationship between mediator and dependent variable has been well established in the literature” (von Soest & Hagtvet, 2011, p. 299).

Several limitations of this study should be noted. First, other previous simulation studies examining the mediated effect have found that the joint significance test is more powerful than the Sobel test (MacKinnon et al., 2002; Taylor et al., 2008; Cheong, 2011), a finding that is consistent with the results of the present study. These are just two of many methods available to test the significance of the mediated effect in addition to a variety of methods to construct confidence intervals (e.g., see MacKinnon et al., 2002; Shrout & Bolger, 2002; MacKinnon et al., 2004; MacKinnon, Fritz, Williams, & Lockwood, 2007). Many of these methods have been assessed in simulation studies of the mediated effect in a number of different study designs. Future studies should

examine the performance of these methods in the two-stage piecewise parallel process model of mediation. Second, the observed variables for M and Y were simulated as continuous. Latent growth curves models can also be used when one has repeated observations of a categorical variable, such as binary, ordinal, or count variables (Muthén, 2004). Sample size requirements may be different from these models, which tend to be more complex. Third, a simplifying assumption of no contemporaneous mediation relationships was made in the present study, which allowed for an exploration of the situation where the true population reflected the strongest case of temporal precedence. It is possible that the presence of such paths may alter the biasing effects associated with model misspecifications. Finally, the putative mediator in this study was the slope (i.e., a growth rate factor). Oftentimes, the intercept factor may be involved in a theoretical model of mediation (Flora et al., 2007; von Soest & Hagtvet, 2011) and future research is needed to assess whether the results of the present study generalize to this setting.

References

- Audrain-McGovern, J., Rodriguez, D., & Kassel, J. D. (2009). Adolescent smoking and depression: Evidence for self-medication and peer smoking mediation. *Addiction*, *104*, 1743-1756.
- Bandalos, D. L. (2006). The use of Monte Carlo studies in structural equation modeling research. In G. R. Hancock & R.O. Mueller (Eds.), *Structural equation modeling: A second course* (pp. 385-426). Greenwich, CT: Information Age Publishing.
- Bollen, K. A., & Curran, P. J. (2006). *Latent curve models: A structural equation perspective*. Hoboken, NJ: Wiley.
- Brown, T. A. (2006). *Confirmatory Factor Analysis for Applied Research*. New York: The Guilford Press.
- Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 136-162). Newbury Park, CA: SAGE.

- Burton, A., Altman, D., Royston, P., & Holder, R. L. (2006). The design of simulations studies in medical statistics. *Statistics in Medicine*, 25, 4279-4292.
- Cerin, E. (2010). Ways of unraveling how and why physical activity influences mental health through statistical mediation analyses. *Mental Health and Physical Activity*, 3, 51-60.
- 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. (2011). Accuracy of estimates and statistical power for testing mediation in latent growth curve modeling. *Structural Equation Modeling*, 18, 195-211.
- Cheung, G. W., & Lau, R. S. (2008). Testing mediation and suppression effects of latent variables: Bootstrapping with structural equation models. *Organizational Research Methods*, 11, 296-325.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*, 2nd ed. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Cohen, J., & Cohen, P. (1983). *Applied multiple regression/correlation analysis for the behavioral sciences*, 2nd ed. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Duncan, T. E., & Duncan, S. C. (2004). An introduction to latent growth curve modeling. *Behavior Therapy*, 35, 333-363.
- Fan, X., & Sivo, S. A. (2005). Sensitivity of fit indexes to misspecified structural or measurement model components: Rationale of two-index strategy revisited. *Structural Equation Modeling*, 12, 343-367.
- Fan, X., & Sivo, S. A. (2007). Sensitivity of fit indices to model misspecification and model types. *Multivariate Behavioral Research*, 42, 509-529.
- Flora, D. B., Khoo, S., & Chassin, L. (2007). Moderating effects of a risk factor: Modeling longitudinal moderated mediation in the development of adolescent heavy drinking. In T. D. Little, J. A. Bovaird, N. A. Card (Eds.), *Modeling contextual effects in longitudinal studies* (pp. 231-254). Mahwah, NJ: Lawrence Erlbaum Associates Publishers.
- Flora, D. B. (2008). Specifying piecewise latent trajectory models for longitudinal data. *Structural Equation Modeling*, 15, 513-533.
- Hancock, G. R., Kuo, W. L., & Lawrence, F. R. (2001). An illustration of second-order latent growth models. *Structural Equation Modeling*, 8, 470-489.
- Heene, M., Hilbert, S., Draxler, C., Ziegler, M., & Bühner, M. (2011). Masking misfit in confirmatory factor analysis by increasing unique variances: A cautionary note on the usefulness of cutoff values of fit indices. *Psychological Methods*, 16, 319-336.
- Hertzog, C., von Oertzen, T., Ghisletta, P., & Lindenberger, U. (2008). Evaluating the power of latent growth curve models to detect individual differences in change. *Structural Equation Modeling*, 15, 541-563.
- Hill, A. (1965). The environment and disease: Association or causation? *Proceedings of the Royal Society of Medicine*, 58, 295-300.
- Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6, 1-55.

- Kaplan, D. (1988). The impact of specification error on the estimation, testing, and improvement of structural equation models. *Multivariate Behavioral Research*, 23, 69-86.
- Kazdin, A. E., & Nock, M. K. (2003). Delineating mechanisms of change in child and adolescent therapy: methodological issues and research recommendations. *Journal of Child Psychology and Psychiatry*, 44, 1116-1129.
- Keppel, G., & Wickens, T. D. (2004). *Design and analysis: A researcher's handbook*, 4th ed. Upper Saddle River, NJ: Pearson Prentice Hall.
- Khoo, S. R. (2001). Assessing program effects in the presence of treatment-baseline interactions: A latent curve approach. *Psychological Methods*, 6, 234-257.
- Kline, R. B. (2011). *Principles and practice of structural equation modeling*, 3rd ed. New York: The Guilford Press.
- Laurenceau, J. P., Hayes, A. M., & Feldman, G. C. (2007). Some methodological and statistical issues in the study of change processes in psychotherapy. *Clinical Psychology Review*, 27, 682-695.
- Lee, I., & Rojewski, J. W. (2009). Development of occupational aspiration prestige: A piecewise latent growth model of selected influences. *Journal of Vocational Behavior*, 75, 82-90.
- Littlefield, A. K., Sher, K. J., & Wood, P. K. (2010). Do changes in drinking motives mediate the relation between personality change and "maturing out" of problem drinking? *Journal of Abnormal Psychology*, 119, 93-105.
- Liu, L. C., Flay, B. R., & Aban Aya Investigators. (2009). Evaluating mediation in longitudinal multivariate data: Mediation effects for the Aban Aya youth project drug prevention program. *Prevention Science*, 10, 197-207.
- 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., Hoffman, J. M., West, S. G., & Sheets, V. (2002). A comparison of methods to test the significance of the mediated effect. *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., 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. (2008a). *Introduction to statistical mediation analysis*. New York: Lawrence Erlbaum Associates.
- MacKinnon, D. P. (2008b). Longitudinal mediation analysis. In D.P. MacKinnon, *Introduction to statistical mediation analysis* (pp. 193-236). New York: Lawrence Erlbaum Associates.
- Marsh, H. W., Hau, K., & Wen, Z. (2004). In search of golden rules: Comment on hypothesis-testing approaches to setting cutoff values for fit indexes and dangers in overgeneralizing Hu and Bentler's (1999) findings. *Structural Equation Modeling*, 11, 320-341.

- Maxwell, S. E., & Cole, D. A. (2007). Bias in cross-sectional analyses of longitudinal mediation. *Psychological Methods*, 12, 23-44.
- McArdle, J. J. (1988). Dynamic but structural equation modeling of repeated measures data. In J. R. Nesselroade, & R. B. Cattell (Eds.), *Handbook of multivariate experimental psychology* (2nd ed., pp. 561-614). New York: Plenum Press.
- Muthén, B. O., & Curran, P. J. (1997). General longitudinal modeling of individual differences in experimental designs: A latent variable framework for analysis and power estimation. *Psychological Methods*, 2, 371-402.
- Muthén, L. K., & Muthén, B. O. (2002). How to use a Monte Carlo study to decide on sample size and determine power. *Structural Equation Modeling*, 9, 599-620.
- Muthén, B. (2004). Latent variable analysis: Growth mixture modeling and related techniques for longitudinal data. In: Kaplan D, ed. *The SAGE handbook of quantitative methodology for the social sciences*. Thousand Oaks, CA: SAGE Publications, 345-368.
- Paxton, P., Curran, P. J., Bollen, K. A., Kirby, J., & Chen F. (2001). Monte Carlo experiments: Design and implementation. *Structural Equation Modeling*, 8, 287-312.
- Raykov, T. (1998). Satisfying a simplex structure is simpler than it should be: A latent curve analysis revisit. *Multivariate Behavior Research*, 33, 343-363.
- Roesch, S. C., Norman, G. J., Adams, M. A., Kerr, J., Sallis, J. F., Ryan, S., & ... Patrick, K. (2009). Latent growth curve modeling of adolescent physical activity: Testing parallel process and mediation models. *Journal of Health Psychology*, 14, 313-325.
- Roesch, S. C., Norman, G. J., Villodas, F., Sallis, J. F., & Patrick, K. (2010). Intervention-mediated effects for adult physical activity: A latent growth curve analysis. *Social Science & Medicine*, 71, 494-501.
- Sayer, A. G., & Willett, J. B. (1998). A cross-domain model for growth in adolescent alcohol expectancies. *Multivariate Behavioral Research*, 33, 509-543.
- Sayer, A. G., & Cumsille, P. E. (2001). Second-order latent growth models. In L. M. Collins & A. G. Sayer (Eds.), *New methods for the analysis of change* (pp. 179-200). Washington, DC: American Psychological Association.
- Selig, J. P., & Preacher, K. J. (2009). Mediation models for longitudinal data in developmental research. *Research in Human Development*, 6, 144-164.
- Shrout, P. E., & Bolger, N. (2002). Mediation in experimental and nonexperimental studies: New procedures and recommendations. *Psychological Methods*, 7, 422-445.
- Singer, J. D., & Willett, J. B. (2003). *Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence*. Oxford: Oxford University Press.
- Sobel, M. E. (1982). Asymptotic confidence intervals for indirect effects in structural equation models. *Sociological Methodology*, 13, 290-312.
- Taylor, A. B., MacKinnon, D. P., Tein J. Y. (2008). Tests of the three-path mediated effect. *Organizational Research Methods*, 11, 241-269.
- Thoemmes, F., MacKinnon, D. P., & Reiser, M. R. (2010). Power analysis for complex mediational designs using Monte Carlo methods. *Structural Equation Modeling*, 17, 510-534.

- von Soest, T., & Hagtvet, K. A. (2011). Mediation analysis in a latent growth curve modeling framework. *Structural Equation Modeling, 18*, 289-314.
- Yuan, K. (2005). Fit indices versus test statistics. *Multivariate Behavioral Research, 40*, 115-148.

Table 1. Model-implied means for the mediator and the outcome under a few simulation conditions

Size of γ_3 and γ_4		Degree of treatment effect on $S2_Y$									
		None ($\beta = 0$ and $\tau' = 0$)					Highest degree in the study ($\beta = 0.65$ and $\tau' = 0.25$)				
		Time 1	Time 2	Time 3	Time 4	Time 5	Time 1	Time 2	Time 3	Time 4	Time 5
$\gamma_3 = 0, \gamma_4 = 0$											
Mediator	Control group	0	0	0	0	0	0	0	0	0	0
	Treatment group	0	0.4	0.8	0.8	0.8	0	0.4	0.8	0.8	0.8
Outcome	Control group	0	0	0	0	0	0	0	0	0	0
	Treatment group	0	0	0	0	0	0	0	0	0.51	1.02
$\gamma_3 = 0.4, \gamma_4 = 0$											
Mediator	Control group	0	0	0	0	0	0	0	0	0	0
	Treatment group	0	0.4	0.8	1.2	1.6	0	0.4	0.8	1.2	1.6
Outcome	Control group	0	0	0	0	0	0	0	0	0	0
	Treatment group	0	0	0	0	0	0	0	0	0.51	1.02
$\gamma_3 = 0, \gamma_4 = 0.4$											
Mediator	Control group	0	0	0	0	0	0	0	0	0	0
	Treatment group	0	0.4	0.8	0.8	0.8	0	0.4	0.8	0.8	0.8
Outcome	Control group	0	0	0	0	0	0	0	0	0	0
	Treatment group	0	0.4	0.8	0.8	0.8	0	0.4	0.8	1.31	1.82
$\gamma_3 = 0.4, \gamma_4 = 0.4$											
Mediator	Control group	0	0	0	0	0	0	0	0	0	0
	Treatment group	0	0.4	0.8	1.2	1.6	0	0.4	0.8	1.2	1.6
Outcome	Control group	0	0	0	0	0	0	0	0	0	0
	Treatment group	0	0.4	0.8	0.8	0.8	0	0.4	0.8	1.31	1.82

Note: $\alpha = 0.4$ under all conditions; $\beta = 0.65$ and $\tau' = 0.25$ are the largest effects used in the study.

Table 2. Mean fit index values across simulation conditions

		Model specification and sample size									
Effect size of mediated effect	Degree of later growth in mediator (path γ_3)	Properly specified two-stage piecewise parallel process model of mediation					Misspecified single-stage parallel process model of mediation				
		$n = 100$	$n = 200$	$n = 500$	$n = 1000$	$n = 2000$	$n = 100$	$n = 200$	$n = 500$	$n = 1000$	$n = 2000$
RMSEA											
Zero	None	0.024	0.016	0.009	0.006	0.004	0.089	0.088	0.088	0.088	0.088
	Moderate	0.026	0.015	0.009	0.006	0.004	0.070	0.068	0.068	0.068	0.068
Small	None	0.024	0.016	0.009	0.006	0.004	0.091	0.090	0.089	0.089	0.089
	Moderate	0.024	0.015	0.009	0.006	0.004	0.070	0.069	0.070	0.070	0.070
Medium	None	0.025	0.016	0.009	0.007	0.004	0.097	0.095	0.095	0.095	0.095
	Moderate	0.024	0.016	0.009	0.006	0.004	0.078	0.077	0.077	0.077	0.077
Large	None	0.024	0.015	0.009	0.006	0.004	0.105	0.105	0.104	0.105	0.104
	Moderate	0.024	0.015	0.009	0.006	0.004	0.089	0.088	0.088	0.088	0.088
CFI											
Zero	None	0.994	0.997	0.999	0.9996	0.9998	0.946	0.949	0.950	0.950	0.950
	Moderate	0.994	0.998	0.999	0.9996	0.9998	0.967	0.971	0.972	0.972	0.972
Small	None	0.994	0.997	0.999	0.9996	0.9998	0.945	0.947	0.948	0.949	0.949
	Moderate	0.995	0.998	0.999	0.9996	0.9998	0.967	0.970	0.970	0.971	0.971
Medium	None	0.994	0.997	0.999	0.9995	0.9998	0.938	0.941	0.942	0.942	0.942
	Moderate	0.995	0.998	0.999	0.9996	0.9998	0.961	0.963	0.964	0.964	0.964
Large	None	0.994	0.998	0.999	0.9996	0.9998	0.927	0.929	0.930	0.930	0.930
	Moderate	0.995	0.998	0.999	0.9996	0.9998	0.950	0.952	0.953	0.953	0.953

Note: Entries are mean fit index values, collapsing across factors for which results were similar (i.e., size of the direct effect and degree of earlier growth in the outcome (path γ_4)). Although sample size or interactions with this factor did not account for nontrivial variance for CFI, it is nevertheless tabled above. Since path α was fixed in this study (0.4), effect size of the mediated effect was defined according the size of path β and reflects zero, small, medium, and large effects on R^2 metric as described by Cohen (1988). A moderate degree of later growth in the mediator was defined as path $\gamma_3 = 0.4$. RMSEA = root mean square error of approximation; CFI = comparative fit index.

Table 3. Mean relative bias of the mediated effect estimates ($\hat{\alpha} * \hat{\beta}$) across simulation conditions

		Model specification and sample size									
Effect size of mediated effect	Degree of later growth in mediator (path γ_3)	Properly specified two-stage piecewise parallel process model of mediation					Misspecified single-stage parallel process model of mediation				
		$n = 100$	$n = 200$	$n = 500$	$n = 1000$	$n = 2000$	$n = 100$	$n = 200$	$n = 500$	$n = 1000$	$n = 2000$
Small	None	0.142	0.225	0.095	0.081	0.027	0.926	0.855	0.447	0.395	0.346
	Moderate	0.265	0.233	0.115	0.052	0.026	2.489	2.032	1.542	1.341	1.294
Medium	None	0.066	0.149	0.071	0.040	0.008	0.025	0.002	-0.143	-0.210	-0.239
	Moderate	0.097	0.163	0.054	0.038	0.020	0.800	0.713	0.424	0.363	0.320
Large	None	-0.024	0.054	0.040	0.031	0.014	-0.141	-0.208	-0.313	-0.353	-0.372
	Moderate	-0.046	0.033	0.049	0.019	0.010	0.387	0.325	0.179	0.113	0.092

Note: Entries are mean relative bias, collapsing across factors for which results were similar (i.e., size of the direct effect and degree of earlier growth in the outcome (path γ_4)). Since path α was fixed in this study (0.4), effect size of the mediated effect was defined according the size of path β and reflects zero, small, medium, and large effects on R^2 metric as described by Cohen (1988). A moderate degree of later growth in the mediator was defined as path $\gamma_3 = 0.4$.

Table 4. Mean bias of the mediated effect estimates ($\hat{\alpha} * \hat{\beta}$) when mediated effect = 0 across simulation conditions

		Model specification and sample size									
Effect size of mediated effect	Degree of later growth in mediator (path γ_3)	Properly specified two-stage piecewise parallel process model of mediation					Misspecified single-stage parallel process model of mediation				
		$n = 100$	$n = 200$	$n = 500$	$n = 1000$	$n = 2000$	$n = 100$	$n = 200$	$n = 500$	$n = 1000$	$n = 2000$
Zero	None	0.016	0.007	0.005	0.002	0.001	0.090	0.080	0.063	0.058	0.054
	Moderate	0.013	0.012	0.004	0.001	0.0002	0.144	0.125	0.103	0.095	0.090

Note: Since the population value is 0, entries are mean bias not relative bias, collapsing across factors for which results were similar (i.e., size of the direct effect and degree of earlier growth in the outcome (path γ_4)). A moderate degree of later growth in the mediator was defined as path $\gamma_3 = 0.4$.

Table 5. Mean relative bias of path $\hat{\beta}$ across simulation conditions

		Model specification and sample size									
Effect size of mediated effect	Degree of later growth in mediator (path γ_3)	Properly specified two-stage piecewise parallel process model of mediation					Misspecified single-stage parallel process model of mediation				
		$n = 100$	$n = 200$	$n = 500$	$n = 1000$	$n = 2000$	$n = 100$	$n = 200$	$n = 500$	$n = 1000$	$n = 2000$
Small	None	0.116	0.240	0.099	0.079	0.027	2.852	2.729	1.878	1.777	1.689
	Moderate	0.254	0.238	0.115	0.047	0.026	2.511	2.037	1.537	1.336	1.289
Medium	None	0.060	0.156	0.069	0.040	0.009	1.096	1.038	0.734	0.595	0.541
	Moderate	0.086	0.158	0.053	0.035	0.017	0.805	0.725	0.437	0.374	0.333
Large	None	-0.027	0.050	0.044	0.032	0.014	0.745	0.612	0.400	0.324	0.281
	Moderate	-0.040	0.039	0.046	0.019	0.010	0.417	0.356	0.201	0.135	0.115

Note: Entries are mean relative bias, collapsing across factors for which results were similar (i.e., size of the direct effect and degree of earlier growth in the outcome (path γ_4)). Since path α was fixed in this study (0.4), effect size of the mediated effect was defined according the size of path β and reflects zero, small, medium, and large effects on R^2 metric as described by Cohen (1988). A moderate degree of later growth in the mediator was defined as path $\gamma_3 = 0.4$.

Table 6. Mean bias of path $\hat{\beta}$ when $\beta = 0$ across simulation conditions

		Model specification and sample size									
Effect size of mediated effect	Degree of later growth in mediator (path γ_3)	Properly specified two-stage piecewise parallel process model of mediation					Misspecified single-stage parallel process model of mediation				
		$n = 100$	$n = 200$	$n = 500$	$n = 1000$	$n = 2000$	$n = 100$	$n = 200$	$n = 500$	$n = 1000$	$n = 2000$
Zero	None	0.035	0.021	0.013	0.006	0.002	0.437	0.402	0.310	0.285	0.265
	Moderate	0.033	0.030	0.011	0.003	0.0005	0.361	0.310	0.255	0.235	0.224

Note: Since the population value is 0, entries are mean bias not relative bias, collapsing across factors for which results were similar (i.e., size of the direct effect and degree of earlier growth in the outcome (path γ_4)). Since path α was fixed in this study (0.4), effect size of the mediated effect was defined according the size of path β . Thus, a zero mediated effect corresponds to $\beta = 0$. A moderate degree of later growth in the mediator was defined as path $\gamma_3 = 0.4$.

Table 7. Mean bias of path $\hat{\tau}'$ across simulation conditions

		Model specification and sample size									
Degree of earlier growth in outcome (path γ_4)	Degree of later growth in mediator (path γ_3)	Properly specified two-stage piecewise parallel process model of mediation					Misspecified single-stage parallel process model of mediation				
		$n = 100$	$n = 200$	$n = 500$	$n = 1000$	$n = 2000$	$n = 100$	$n = 200$	$n = 500$	$n = 1000$	$n = 2000$
Complete mediation ($\tau' = 0$)											
None	None	-0.006	-0.012	-0.008	-0.005	-0.002	-0.086	-0.074	-0.051	-0.041	-0.037
	Moderate	-0.005	-0.019	-0.007	-0.006	-0.002	-0.191	-0.178	-0.140	-0.127	-0.122
Moderate	None	-0.011	-0.014	-0.009	-0.002	-0.001	0.114	0.133	0.154	0.160	0.164
	Moderate	-0.001	-0.024	-0.008	-0.005	-0.004	0.005	0.027	0.062	0.074	0.077
Partial mediation ($\tau' = 0.25$)											
None	None	-0.013	-0.019	-0.007	-0.007	-0.002	-0.218	-0.212	-0.180	-0.171	-0.168
	Moderate	-0.011	-0.013	-0.007	-0.001	-0.001	-0.325	-0.305	-0.272	-0.259	-0.255
Moderate	None	-0.001	-0.015	-0.006	-0.007	-0.002	0.003	0.003	0.028	0.034	0.040
	Moderate	-0.007	-0.013	-0.011	-0.005	-0.002	-0.110	-0.094	-0.062	-0.051	-0.044

Note: Entries are mean bias, collapsing across factors for which results were similar (i.e., size of the mediated effect). A moderate degree of later growth in the mediator was defined as path $\gamma_3 = 0.4$ and a moderate degree of earlier growth in the outcome was defined as path $\gamma_4 = 0.4$. Although it is possible to calculate relative bias when $\tau' = 0.25$, for consistency, the values are mean bias not relative bias.

Table 8. Mean relative bias of standard error of the mediated effect ($\hat{\alpha} * \hat{\beta}$) across simulation conditions

Effect size of mediated effect	Model specification and sample size									
	Properly specified two-stage piecewise parallel process model of mediation					Misspecified single-stage parallel process model of mediation				
	$n = 100$	$n = 200$	$n = 500$	$n = 1000$	$n = 2000$	$n = 100$	$n = 200$	$n = 500$	$n = 1000$	$n = 2000$
Zero	0.317	0.012	-0.001	-0.015	-0.003	0.449	0.041	-0.092	-0.018	-0.019
Small	0.286	0.091	-0.006	0.001	-0.005	0.352	0.003	-0.031	-0.034	-0.007
Medium	0.403	0.105	-0.017	-0.011	0.003	0.425	0.031	-0.028	-0.007	-0.007
Large	0.430	0.149	0.013	0.015	0.004	0.412	0.053	-0.059	-0.031	-0.007

Note: Entries are mean relative bias, collapsing across factors for which results were similar (i.e., size of the direct effect, degree of later growth in the mediator (path γ_3), and degree of earlier growth in the outcome (path γ_4)). Since path α was fixed in this study (0.4), effect size of the mediated effect was defined according the size of path β and reflects zero, small, medium, and large effects on R^2 metric as described by Cohen (1988). Standard error of the mediated effect was obtained using Sobel's (1982) method.

Table 9. Estimated Type I error rates for the mediated effect across simulation conditions

Method for testing mediated effect	Model specification and sample size									
	Properly specified two-stage piecewise parallel process model of mediation					Misspecified single-stage parallel process model of mediation				
	$n = 100$	$n = 200$	$n = 500$	$n = 1000$	$n = 2000$	$n = 100$	$n = 200$	$n = 500$	$n = 1000$	$n = 2000$
Sobel test	0.006	0.018	0.034	0.045	0.048	0.002	0.011	0.141	0.434	0.773
Joint significance test	0.015	0.028	0.038	0.048	0.049	0.004	0.014	0.152	0.442	0.775

Note: Entries are mean Type I error rates for the mediated effect, collapsing across factors for which results were similar (i.e., size of the direct effect, degree of later growth in the mediator (path γ_3), and degree of earlier growth in the outcome (path γ_4)).

Table 10. Empirical power levels across simulation conditions for the test of the mediated effect

Effect size for mediated effect	Model specification and sample size									
	Properly specified two-stage piecewise parallel process model of mediation					Misspecified single-stage parallel process model of mediation				
	$n = 100$	$n = 200$	$n = 500$	$n = 1000$	$n = 2000$	$n = 100$	$n = 200$	$n = 500$	$n = 1000$	$n = 2000$
	Sobel test									
Small	0.006	0.024	0.145	0.324	0.606	0.002	0.027	0.339	0.794	0.988
Medium	0.022	0.167	0.717	0.970	1.000	0.010	0.094	0.707	0.983	1.000
Large	0.065	0.420	0.968	1.000	1.000	0.024	0.188	0.880	1.000	1.000
	Joint significance test									
Small	0.017	0.036	0.156	0.335	0.610	0.005	0.035	0.357	0.798	0.988
Medium	0.050	0.216	0.739	0.973	1.000	0.017	0.118	0.727	0.984	1.000
Large	0.134	0.497	0.971	1.000	1.000	0.040	0.228	0.889	1.000	1.000

Note: Entries are mean power levels, collapsing across factors for which results were similar (i.e., size of the direct effect, degree of later growth in the mediator (path γ_3), and degree of earlier growth in the outcome (path γ_4)). Since path α was fixed in this study (0.4), effect size of the mediated effect was defined according the size of path β and reflects small, medium, and large effects on R^2 metric as described by Cohen (1988).

Table 11. Estimated Type I error rates and empirical power for the direct effects across simulation conditions levels

		Model for the mediator and sample size									
Degree of earlier growth in outcome (path γ_4)	Degree of later growth in mediator (path γ_3)	Properly specified two-stage piecewise parallel process model of mediation					Misspecified single-stage parallel process model of mediation				
		$n = 100$	$n = 200$	$n = 500$	$n = 1000$	$n = 2000$	$n = 100$	$n = 200$	$n = 500$	$n = 1000$	$n = 2000$
Type I error – Complete mediation ($\tau' = 0$)											
None	None	0.025	0.036	0.042	0.043	0.047	0.020	0.014	0.026	0.084	0.202
	Moderate	0.015	0.021	0.034	0.044	0.051	0.006	0.022	0.160	0.497	0.830
Moderate	None	0.025	0.036	0.030	0.047	0.051	0.338	0.470	0.694	0.892	0.986
	Moderate	0.018	0.023	0.043	0.051	0.047	0.119	0.147	0.228	0.335	0.486
Power – Partial mediation ($\tau' = 0.25$)											
None	None	0.277	0.478	0.822	0.966	0.999	0.169	0.214	0.353	0.524	0.757
	Moderate	0.149	0.268	0.590	0.864	0.981	0.046	0.046	0.045	0.069	0.110
Moderate	None	0.299	0.482	0.824	0.957	0.999	0.590	0.701	0.897	0.990	1.000
	Moderate	0.169	0.299	0.582	0.851	0.983	0.306	0.394	0.599	0.802	0.950

Note: Entries are mean Type I error rate (when $\tau' = 0$) and power (when $\tau' = 0.25$), collapsing across factors for which results were similar (i.e., size of the mediated effect). A moderate degree of later growth in the mediator was defined as path $\gamma_3 = 0.4$ and a moderate degree of earlier growth in the outcome was defined as path $\gamma_4 = 0.4$.

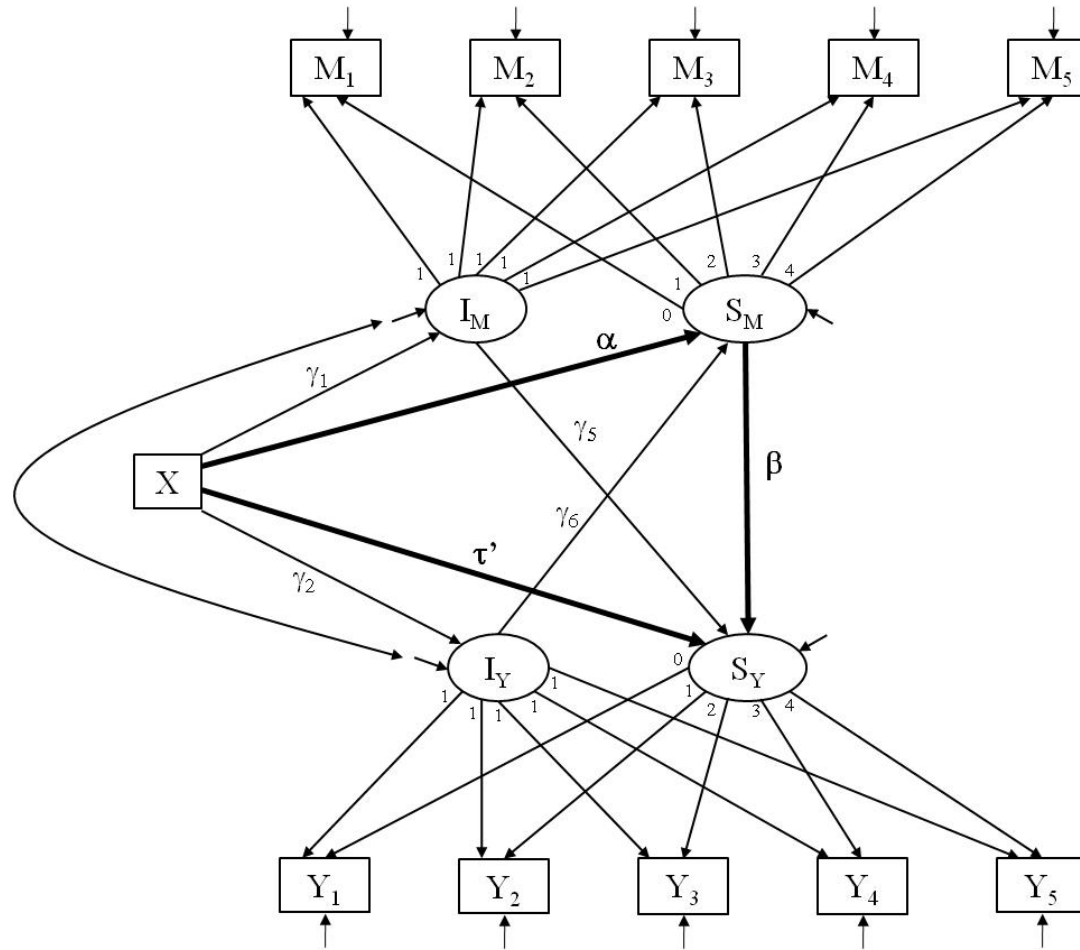


Figure 1. Parallel process latent growth curve mediation model with binary X variable (the independent variable) and five waves of data for the mediator (M) and the dependent variable (Y) (adapted from Cheong et al., 2003). I_M represents the intercept factor (i.e., initial status) of the mediator, S_M represents the slope (growth rate) factor of the mediator, I_Y represents the intercept factor (i.e., initial status) of the outcome, and S_Y represents the slope (growth rate) factor of the outcome. Note that correlations between growth factors (e.g., between initial status and disturbances of slope factor for M) are not shown to simplify the figure.

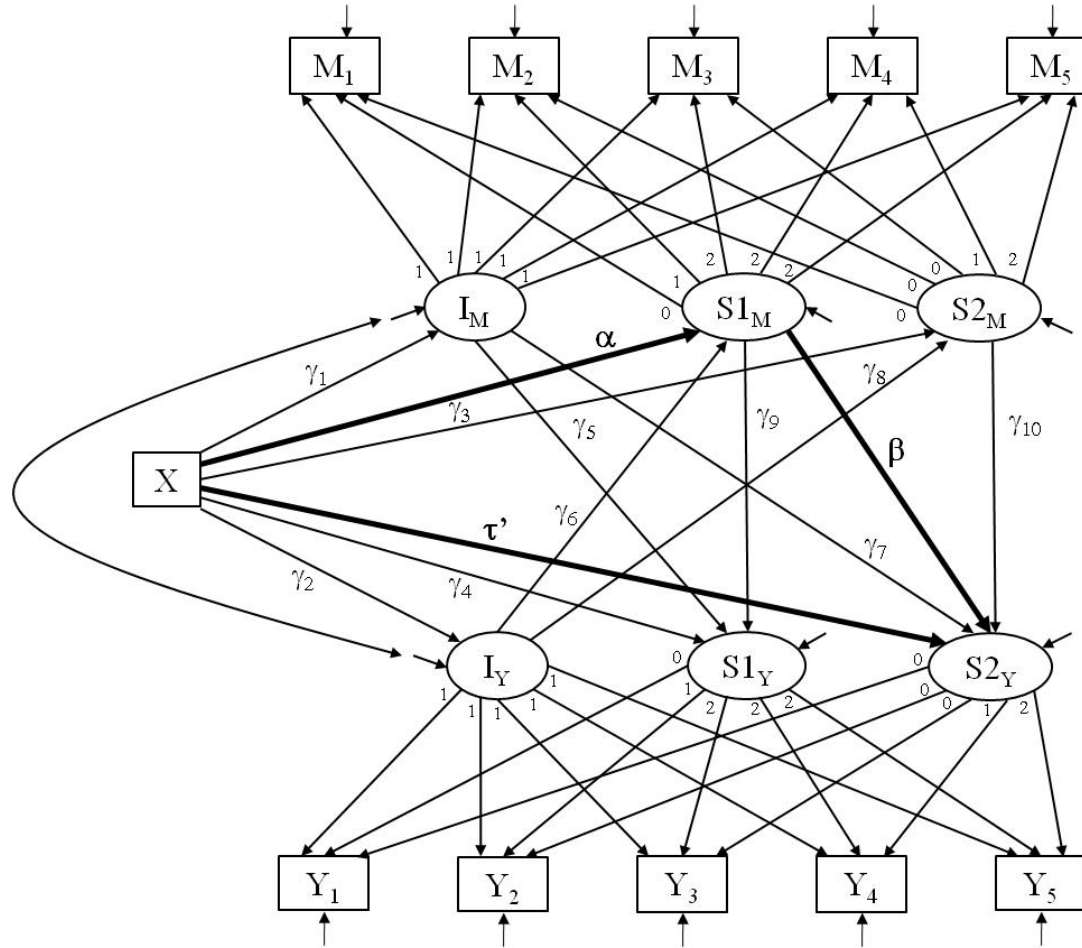


Figure 2. Two-stage piecewise parallel process latent growth curve mediation model with binary X variable (the independent variable) and five waves of data for the mediator (M) and the dependent variable (Y). I_M and I_Y represent initial status (i.e., at baseline) for the mediator and the outcome, respectively. $S1_M$ and $S1_Y$ represent growth rates for the mediator and outcome, respectively, at an earlier stage of the study; and $S2_M$ and $S2_Y$, represent growth rates for the mediator and outcome, respectively, at a later stage of the study. Note that correlations between growth factors (e.g., between initial status and disturbances of the growth rate factor for M) are not shown to simplify the figure.

Appendix – Mplus Source Code

!Mplus data generation code for true population model where size of the
!mediated effect = medium (beta = 0.40), partial mediation (tau-prime =
!0.25), no late growth in the mediator (gamma3 = 0), no early growth in
!the outcome (gamma4 = 0), and sample size=1000

MONTECARLO:

```
names = m1 m2 m3 m4 m5 y1 y2 y3 y4 y5 x;  
cutpoints=x(0);  
nobs = 1000;  
nreps = 1000;  
seed = 9516995;  
repsave = all;  
save = cell_124rep*.dat;
```

MODEL MONTECARLO:

```
[x@0]; x@1;  
im s1m | m1@0 m2@1 m3@2 m4@2 m5@2;  
im s2m | m1@0 m2@0 m3@0 m4@1 m5@2;  
m1-m5@0.25;  
[im@0 s1m@0 s2m@0];  
im@1; s1m@0.1; s2m@0.1;  
im with s1m@-0.1;  
im with s2m@-0.05;  
s1m with s2m@-0.05;  
iy s1y | y1@0 y2@1 y3@2 y4@2 y5@2;  
iy s2y | y1@0 y2@0 y3@0 y4@1 y5@2;  
y1-y5@0.25;  
[iy@0 s1y@0 s2y@0];  
iy@1; s1y@0.1; s2y@0.1;  
iy with s1y@-0.1;  
iy with s2y@-0.05;  
s1y with s2y@-0.05;  
im with iy@0.3;  
im on x@0; !gamma1  
s1m on iy@-0.1; !gamma6  
s1m on x@0.4; !alpha  
s2m on iy@-0.1; !gamma8  
s2m on x@0; !gamma3  
iy on x@0; !gamma2  
s1y on im@-0.1; !gamma5  
s1y on s1m@0; !gamma9  
s1y on x@0; !gamma4  
s2y on im@-0.1; !gamma7  
s2y on s1m@0.4; !beta  
s2y on x@0.25; !tau-prime;  
s2y on s2m@0; !gamma10
```

```
!Mplus code that reads in the data created from the first syntax and
!fits the properly specified two-stage piecewise parallel process model
!of mediation
```

```
DATA:    FILE=cell_124replist.dat;
        TYPE = MONTECARLO;
VARIABLE: NAMES = m1 m2 m3 m4 m5 y1 y2 y3 y4 y5 x;
        USEVARIABLES = ALL;
MODEL:
    im s1m | m1@0 m2@1 m3@2 m4@2 m5@2;
    im s2m | m1@0 m2@0 m3@0 m4@1 m5@2;
    m1-m5*0.25;
    [im*0 s1m*0 s2m*0];
    im*1; s1m*0.1; s2m*0.1;
    im with s1m*-0.1;
    im with s2m*-0.05;
    s1m with s2m*-0.05;
    iy s1y | y1@0 y2@1 y3@2 y4@2 y5@2;
    iy s2y | y1@0 y2@0 y3@0 y4@1 y5@2;
    y1-y5*0.25;
    [iy*0 s1y*0 s2y*0];
    iy*1; s1y*0.1; s2y*0.1;
    iy with s1y*-0.1;
    iy with s2y*-0.05;
    s1y with s2y*-0.05;
    im with iy*0.3;
    im on x*0; !gamma1
    s1m on iy*-0.1; !gamma6
    s1m on x*0.4 (a); !alpha
    s2m on iy*-0.1; !gamma8
    s2m on x*0; !gamma3
    iy on x*0; !gamma2
    s1y on im*-0.1; !gamma5
    s1y on s1m*0; !gamma9
    s1y on x*0; !gamma4
    s2y on im*-0.1; !gamma7
    s2y on s1m*0.4 (b); !beta
    s2y on x*0.25; !tau-prime;
    s2y on s2m*0; !gamma10
MODEL CONSTRAINT:
    NEW(med*0.16);
    med=a*b;
OUTPUT: TECH9;
SAVEDATA: RESULTS=cell_124.txt;
```

```

!Mplus code that reads in the data created from the first syntax and
!fits the misspecified single-stage parallel process model of mediation

DATA:  FILE=cell_124replist.dat;
      TYPE = MONTECARLO;
VARIABLE:  NAMES = m1 m2 m3 m4 m5 y1 y2 y3 y4 y5 x;
          USEVARIABLES = ALL;
MODEL:
    im sm | m1@0 m2@1 m3@2 m4@3 m5@4;
    m1-m5*0.25;
    [im*0 sm*0];
    im*1; sm*0.1;
    im with sm*-0.1;
    iy sy | y1@0 y2@1 y3@2 y4@3 y5@4;
    y1-y5*0.25;
    [iy*0 sy*0];
    iy*1; sy*0.1;
    iy with sy*-0.1;
    im with iy*0.3;
    im on x*0; !gamma1
    sm on iy*-0.1; !gamma6
    sm on x*0.4 (a); !alpha
    iy on x*0; !gamma2
    sy on im*-0.1; !gamma5
    sy on sm*0.4 (b); !beta
    sy on x*0.25; !tau-prime;
MODEL CONSTRAINT:
    NEW(med*0.16);
    med=a*b;
OUTPUT: TECH9;
SAVEDATA: RESULTS=cell_284.txt;

```

SUMMARY, CONCLUSIONS, AND FUTURE RESEARCH

Mediation modeling is commonly used in a number of different disciplines to answer questions about how or why one variable exerts its influence on another variable. Although mediation can be assessed in the context of several different types of study designs, the use of cross-sectional data and a single-mediator model (often with continuous measures of the independent variable, mediator, and dependent variable) is still the norm in most empirical tests of mediation. There are several limitations associated with assessing mediation with cross-sectional data, perhaps the most significant is that mediated effect estimates are biased in the case of true longitudinal mediation (Maxwell & Cole, 2007).

There are several classes of models for evaluating longitudinal mediation with the collection of three or more waves of data. These models are increasingly utilized in the applied literature and methodological research continues to evaluate them, as well as extensions and new approaches. However, despite their use in substantive research, the preponderance of mediation hypotheses are still tested with cross-sectional data. Furthermore, consensus on the optimal implementation of these longitudinal mediation modeling methods is largely lacking and there are still many unanswered questions.

This dissertation sought to demonstrate the application of one approach to longitudinal mediation modeling, namely the autoregressive model, and build on a set of steps recommended for testing such models (paper one). In addition, using two different

simulation studies, we attempted to address significant methodological questions related to both the autoregressive mediation model (paper two) as well as the two-stage piecewise parallel process latent growth curve model of mediation (paper three). The latter model, while mentioned in the literature as a possible method to address a substantial criticism associated with single-stage parallel process mediation models (i.e., the inability to delineate temporal ordering in such models), is rarely utilized in substantive research and lacks a full elaboration in the literature as well as an evaluation of its statistical performance. Below we give a summary of the findings from each paper, discuss some general avenues of future research in longitudinal mediation models, and provide a few implications for researchers wishing to test mediational hypotheses with longitudinal data.

5.1 Concluding Comments on Paper One

In paper one, based on the Wilson-Cleary (1995) conceptual model of patient outcomes, we proposed that the relationship between functional status (disability) and health-related quality of life is mediated by life-space mobility and tested this hypothesis using an autoregressive mediation model based on four waves of data. A set of steps outlined by Cole and Maxwell (2003) and MacKinnon (2008b) for autoregressive mediation modeling was modified and used in this study. Some previous evaluations of the Wilson-Cleary model have used longitudinal data, and have even used autoregressive modeling (Mathisen et al., 2007), but these studies did not test for mediated effects. To our knowledge, this is the first test of mediated relationships implied by the Wilson-Cleary model to use longitudinal data with three or more waves of data. Results from the

autoregressive mediation models supported the mediating role of life-space mobility and suggests that this role is more significant with the mental component summary score from the SF-12 as the outcome compared to the physical component summary score. Model modifications guided by theory and empirical findings did not alter the substantive meaning of this mediated effect, only enhanced it. Estimates of the mediation parameters from the autoregressive models were only partially consistent with mediation estimates from the cross-sectional analyses, suggesting that mediating relationships would have been missed or were potentially underestimated in the cross-sectional models. Flexibility associated with autoregressive modeling allowed for an examination of cross-lagged relationships. This analysis, based on modification indices, suggested the possibility of a reciprocal relationship over time between disability and life-space mobility. Other evaluations of the Wilson-Cleary have examined such reciprocal effects. For example, Mathisen and colleagues (2007) note reciprocal causal effects over time between general health perceptions and overall quality of life. Based paper one, it appears that a framework based on an autoregressive model is useful for exploring mediation relationships.

5.2 Concluding Comments on Paper Two

The proposed mediator in paper one, life-space mobility, was highly reliable and was constructed to be a composite variable. However, there are times when researchers use measures that are less reliable and also have the opportunity to extract latent variables to test a mediational model, rather than simply construct a linear composite based on a set of fallible indicators. In addition, in the context of an autoregressive mediation model,

where variables are measured at several different time points, the possibilities of model misspecification, including the failure to account for shared method variance (e.g., the method effects associated with repeated administrations of the same measure), increases.

The purpose of paper two was to explore the effects of these variables (i.e., the method to handle fallible indicators, omission of paths representing shared method variance). A simulation study was conducted to assess the impact on estimation as well as statistical power and Type I error rates of failing to account for random measurement error and shared method variance in the mediator under a variety of conditions, including the degree of shared method variance, degree of composite reliability, degree of stability of the latent mediator, size of the mediated effect, and sample size. The results from paper two demonstrate that failure to account for measurement error and shared method variance can have a significant impact on parameter estimation in the autoregressive mediation model, including both overestimation and underestimation of paths of interest. Although the extraction of latent variables from multiple observed measures generally provides accurate estimates and also allows researchers to take into account method effects by allowing correlated measurement errors, latent variable models still require significant levels of composite reliability to achieve acceptable levels of power to detect the mediated effect.

5.3 Concluding Comments on Paper Three

The autoregressive mediation model applied in paper one and evaluated in paper two is one of several classes of models for evaluating longitudinal mediation with the collection of three or more waves of data. Latent growth curve modeling can also be

used to assess longitudinal mediation, with one such approach being the parallel process model of mediation. Although this approach has several advantages over cross-sectional tests of mediation, a common criticism is that the model cannot be used to establish that prior changes in the mediator are related to future changes in the outcome (i.e., a lack of temporal precedence). Although definitive cause and effect statements still cannot be made about the mediator-outcome relationship, the two-stage piecewise parallel process model of mediation can be used to test whether early growth in a mediator is related to later growth in an outcome (i.e., temporal sequentiality can be established).

Paper three provides an overview of the two-stage piecewise parallel process model of mediation and reports on a simulation study designed to examine the statistical performance of methods used to test mediation in such a model and also to examine the impact of misspecifying a true piecewise model as a single-stage parallel process model of mediation that assumes linear growth trajectories under a variety of conditions. Results demonstrate that fairly large samples, in some cases 1,000 or more, were generally required to minimize bias of mediated effect estimates and to achieve adequate statistical power. Furthermore, under the conditions of the present study, LGC models of mediation are quite sensitive to model misspecifications that fail to account for the true state of temporal precedence, both in terms of model misfit and parameter estimate bias, suggesting that caution should be exercised in the interpretation of single-stage parallel process mediation models without strong theory linking growth in the mediator and growth in the outcome or without prior established evidence of the temporal relationship between mediator and outcome. Although the two-stage piecewise parallel process model of mediation can be very useful and provides convincing tests of mediational

hypotheses, it is important to note that it requires designs that have at least five waves of data, a constraint that may limit its use in applied studies given limited resources and the logistical difficulties associated with longitudinal studies (von Soest & Hagtvet, 2011).

5.4 Future Research Directions

In addition to some of the specific future areas of research mentioned in the discussion sections of the individual papers, there are several general future avenues of research with respect to longitudinal mediation modeling that require exploration. First, although there are some guidelines and recommendations for selecting which class of models is most appropriate for longitudinal mediation modeling (MacKinnon, 2008b; Selig & Preacher, 2009; Cheong, 2011; von Soest & Hagtvet, 2011), there is limited research to support such recommendations. Second, it is possible to combine features from the multiple classes of longitudinal mediation models. For example, autoregressive latent trajectory (ALT) models, described by Curran and Bollen (2001), combine elements of autoregressive models and latent growth curve models. Selig and Preacher (2009) also briefly mention hybrid models that combine autoregressive and latent growth curve models. Future study directed at such combinations and their statistical performance is warranted. Third, the classes of longitudinal mediation models applied and evaluated in this dissertation, and the additional classes briefly mentioned in Chapter 1, do not comprise an exhaustive list. Future work focused on the development of new methods is also necessary.

The fourth general avenue of future research is the application of existing classes, and potentially the development of new models, for longitudinal mediation modeling

with categorical data, including variables considered to be binary, nominal, ordinal, or count variables. Some work has addressed the assessment of mediation with categorical variables (Huang, Sivaganesan, Succop, & Goodman, 2004; MacKinnon et al., 2007; MacKinnon, 2008f), but this work has focused on in assessment of mediation in the context of cross-sectional designs. In addition to these types of variables, some variables that may be involved in mediational relationships are measured with a preponderance of zeros (i.e., clumping at zero). Because longitudinal mediation models have been primarily developed and tested with the assumption of continuous variables from a multivariate normal distribution, these models may not generalize to situations where the independent variable, mediator, and/or outcome variable have significant clumping at zero. Finally, it is important to understand when cross-sectional analyses will lead to biased estimates of true longitudinal mediation effects. Maxwell and Cole (2007) nicely demonstrate how the pattern of bias depends of different factors for the autoregressive mediation model and one type of random effects model of mediation. Similar analyses are needed for other longitudinal mediation models.

5.5 Implications for Practice

The results of this dissertation suggest that caution should be exercised when interpreting cross-sectional models of mediation. In general, the findings also encourage the appropriate use of longitudinal mediation modeling, building on the efforts of several others who have encouraged the use of these models over the past 20 years. However, care should be taken not to overstep in the interpretation of certain longitudinal models, namely single-stage parallel process models of mediation and autoregressive models that

utilize measures with low reliability and fail to at least discuss the possibility of shared method variance.

Random measurement error is a critical design feature in longitudinal studies and is not sufficient to simply obtain multiple measures and extract latent variables. While such methods may correct the biasing effects of measurement error, they usually will not enhance statistical power, as such models tend to be more complex with more parameters to estimate. The use of highly reliable data, with or without the extraction of latent variables, is important for longitudinal models of mediation. Two major barriers to the use of longitudinal mediation models are logistics and cost, because of the need to follow up (sometimes) large numbers of individuals across three or more measurement waves. As discussed in paper three, piecewise models need data from at least five time points. Thus, any design feature, such as the use of highly reliable measures, which allows for the use of fewer subjects while maintaining appropriate levels of power can help to address logistical and cost-related barriers to the use of longitudinal studies to assess mediation. Researchers are also encouraged to conduct power analysis using Monte Carlo methods to determine sample size prior to conducting a longitudinal study aimed at assessing mediation (Muthén & Muthén, 2002; Thoemmes, MacKinnon, & Reiser, 2010).

GENERAL LIST OF REFERENCES

- Bandalos, D. L. (2006). The use of Monte Carlo studies in structural equation modeling research. In G. R. Hancock & R. O. Mueller (Eds.), *Structural equation modeling: A second course* (pp. 385-426). Greenwich, CT: Information Age Publishing.
- 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.
- Bedeian, A. G., Day, D. V., & Kelloway, E. K. (1997). Correcting for measurement error attenuation in structural equation models: Some important reminders. *Educational and Psychological Measurement*, 57, 785-799.
- Bollen, K. A. (1987). Total, direct, and indirect effects in structural equation models. *Sociological Methodology*, 17, 37-69.
- Bollen, K. A. (2002). Latent variables in psychology and the social sciences. *Annual Review of Psychology*, 53, 605-634.
- Brown, T. A. (2006). *Confirmatory Factor Analysis for Applied Research*. New York: The Guilford Press.
- Cerin, E. (2010). Ways of unraveling how and why physical activity influences mental health through statistical mediation analyses. *Mental Health and Physical Activity*, 3, 51-60.
- 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. (2011). Accuracy of estimates and statistical power for testing mediation in latent growth curve modeling. *Structural Equation Modeling*, 18, 195-211.
- Cheung, G. W., & Lau, R. S. (2008). Testing mediation and suppression effects of latent variables: Bootstrapping with structural equation models. *Organizational Research Methods*, 11, 296-325.
- Cheung, M. W. L. (2007). Comparison of approaches to constructing confidence intervals for mediating effects using structural equation models. *Structural Equation Modeling*, 14, 227-246.
- Cheung, M. W. L. (2009). Comparison of methods for constructing confidence intervals of standardized indirect effects. *Behavior Research Methods*, 41, 425-438.
- Cohen, J., & Cohen, P. (1983). *Applied multiple regression/correlation analysis for the behavioral sciences*, 2nd ed. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Cole, D. A., & Maxwell, S. E. (2003). Testing mediation models with longitudinal data: Questions and tips in the use of structural equation modeling. *Journal of Abnormal Psychology*, 112, 558-577.
- Cole, D. A., Ciesla, J. A., & Steiger, J. H. (2007). The insidious effects of failing to include design-driven correlated residuals in latent-variable covariance structure analysis. *Psychological Methods*, 12, 381-398.
- Collins, L. M., Graham, J. W., & Flaherty, B. P. (1998). An alternative framework for defining mediation. *Multivariate Behavioral Research*, 33, 295-312.

- Collins, L. M., & Graham, J. W. (2002). The effect of the timing and spacing of observations in longitudinal studies of tobacco and other drug use: temporal design considerations. *Drug and Alcohol Dependence*, 68, S85-S96.
- Curran, P. J., & Bollen, K. A. (2001). The best of both worlds: Combining autoregressive and latent curve models. In L. M. Collins & A. G. Sayer (Eds.), *New methods for the analysis of change* (pp. 107-135). Washington, DC: American Psychological Association.
- Dow, W. H., & Norton, E. C. (2003). Choosing between and interpreting the Heckit and two-part models for corner solutions. *Health Services & Outcomes Research Methodology*, 4, 5-18.
- Duncan, T. E., & Duncan, S. C. (2004). An introduction to latent growth curve modeling. *Behavior Therapy*, 35, 333-363.
- 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.
- Fairchild, A. J., & MacKinnon, D. P. (2009). A general model for testing mediation and moderation effects. *Prevention Science*, 10, 87-99.
- Fan, X., & Fan, X. (2005). Using SAS for Monte Carlo Simulation research in SEM. *Structural Equation Modeling*, 12, 299-333.
- Feng, Z., Diehr, P., Peterson, A., & McLerran, D. (2001). Selected statistical issues in group randomized trials. *Annual Review of Public Health*, 22, 167-187.
- 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.
- Freedman, L. S., Graubard, B. I., & Schatzkin, A. (1992). Statistical validation of intermediate endpoints for chronic diseases. *Statistics in Medicine*, 11, 167-178.
- Fritz, M. S., & MacKinnon, D. P. (2007). Required sample size to detect the mediated effect. *Psychological Science*, 18, 233-239.
- Gao, F., Thompson, P., Xiong, C., & Miller, J. P. (2006). Analyzing multivariate longitudinal data using SAS®. SUGI 31 Proceedings. Retrieved from <http://www2.sas.com/proceedings/sugi31/187-31.pdf>.
- Geiser, C. G., Eid, M., Nussbeck, F. W., Courvoisier, D. S., & Cole, D. A. (2010). Analyzing true change in longitudinal multitrait-multimethod studies: Application of a multimethod change model to depression and anxiety in children. *Developmental Psychological*, 46, 29-45.
- Gerbing, D. W., & Anderson, J. C. (1993). Monte Carlo evaluations of goodness-of-fit indices for structural equation models. In K. A. Bollen, & J. S. Long (Eds.), *Testing structural equation models* (pp. 40-65). Newbury Park, CA: Sage.
- 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.

- Guttman, L. A. (1954). A new approach to factor analysis: The radex. In P. F. Lazarfeld (Ed.), *Mathematical thinking in the social sciences* (pp. 258-348). New York: Columbia University Press.
- Hancock, G. R., Kuo, W. L., & Lawrence, F. R. (2001). An illustration of second-order latent growth models. *Structural Equation Modeling*, 8, 470-489.
- Hill, A. (1965). The environment and disease: Association or causation? *Proceedings of the Royal Society of Medicine*, 58, 295-300.
- Hox, J., & Stoel, R. D. (2005). Multilevel and SEM approaches to growth curve modeling. In B. S. Everitt, & D. C. Howell (Eds.), *Encyclopedia of statistics in behavioral science* (pp. 1296-1305). Chichester: John Wiley & Sons.
- Hoyle, R. H., & Kenny, D. A. (1999). Sample size, reliability, and tests of statistical mediation. In R. H. Hoyle (Ed.), *Statistical strategies for small sample research* (pp. 195-222). Thousand Oaks, CA: Sage.
- Huang, B., Sivaganesan, S., Succop, P., & Goodman, E. (2004). Statistical assessment of mediational effects for logistic mediational models. *Statistics in Medicine*, 23, 2713-2728.
- Judd, C. M., & Kenny, D. A. (1981). Process analysis: Estimating mediation in treatment evaluations. *Evaluation Review*, 5, 602-619.
- Kaplan, D. (2009). *Structural Equation Modeling: Foundations and Extensions*, 2nd ed. Los Angeles: Sage.
- Kazdin, A. E., & Nock, M. K. (2003). Delineating mechanisms of change in child and adolescent therapy: methodological issues and research recommendations. *Journal of Child Psychology and Psychiatry*, 44, 1116-1129.
- Kennedy, P. (2003). *A Guide to Econometrics*, 5th ed. Cambridge, MA: The MIT Press.
- Kenny, D. A. (1979). *Correlation and Causality*. New York: Wiley.
- Kenny, D. A., Korchmaros, J. D., & Bolger N. (2003). Lower level mediation in multilevel models. *Psychological Methods*, 8, 115-128.
- Kline, R. B. (2011). *Principles and practice of structural equation modeling*, 3rd ed. New York: The Guilford Press.
- Krull, J. L., & MacKinnon, D. P. (2001). Multilevel modeling of individual and group level mediated effects. *Multivariate Behavioral Research*, 36, 249-277.
- Laurenceau, J. P., Hayes, A. M., & Feldman, G. C. (2007). Some methodological and statistical issues in the study of change processes in psychotherapy. *Clinical Psychology Review*, 27, 682-695.
- Little, T. D., Preacher, K. J., Selig, J. P., & Card, N. A. (2007). New developments in latent variable panel analyses of longitudinal data. *International Journal of Behavioral Development*, 31, 357-365.
- Liu, L. C., Flay, B. R., & Aban Aya Investigators. (2009). Evaluating mediation in longitudinal multivariate data: Mediation effects for the Aban Aya youth project drug prevention program. *Prevention Science*, 10, 197-207.
- MacKinnon, D. P., & Dwyer, J. H. (1993). Estimating mediated effects in prevention studies. *Evaluation Review*, 17, 144-158.
- MacKinnon, D. P., Warsi, G., & Dwyer, J. H. (1995). A simulation study of mediated effect measures. *Multivariate Behavioral Research*, 30, 41-62.
- 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., Hoffman, J. M., West, S. G., & Sheets, V. (2002). A comparison of methods to test the significance of the mediated effect. *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., Lockwood, C. M., Brown, C. H., Wang, W., & Hoffman, J. M. (2007). The intermediate endpoint effect in logistic and probit regression. *Clinical Trials*, 4, 499-513.
- 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., Fairchild, A. J., & Fritz, M. S. (2007). Mediation analysis. *Annual Review of Psychology*, 58, 593-614.
- MacKinnon, D. P. (2008a). *Introduction to statistical mediation analysis*. New York: Lawrence Erlbaum Associates.
- MacKinnon, D. P. (2008b). Longitudinal mediation analysis. In D. P. MacKinnon, *Introduction to statistical mediation analysis* (pp. 193-236). New York: Lawrence Erlbaum Associates.
- MacKinnon, D. P. (2008c). Multiple mediator model. In D. P. MacKinnon, *Introduction to statistical mediation analysis* (pp. 103-125). New York: Lawrence Erlbaum Associates.
- MacKinnon, D. P. (2008d). Causal inference for mediation models. In D. P. MacKinnon, *Introduction to statistical mediation analysis* (pp. 347-367). New York: Lawrence Erlbaum Associates.
- MacKinnon, D. P. (2008e). Multilevel mediation models. In D. P. MacKinnon, *Introduction to statistical mediation analysis* (pp. 237-274). New York: Lawrence Erlbaum Associates.
- MacKinnon, D. P. (2008f). Mediation in categorical data analysis. In D. P. MacKinnon, *Introduction to statistical mediation analysis* (pp. 297-323). New York: Lawrence Erlbaum Associates.
- Mandys, F., Dolan, C. V., & Molenaar, P. C. M. (1994). Two aspects of the simplex model : Goodness of fit to linear growth curve structures and the analysis of mean trends. *Journal of Educational and Behavioral Statistics*, 19, 201-215.
- Marsh, H. W. (1993). Stability of individual differences in multiwave panel studies: Comparison of simplex models and one-factor models. *Journal of Educational Measurement*, 30, 157-183.
- Mathisen, L., Andersen, M. H., Veenstra, M., Wahl, A. K., Hanestad, B. R., & Fosse, E. (2007). Quality of life can both influence and be an outcome of general health perceptions after heart surgery. *Health and Quality of Life Outcomes*, 5, 27.
- Maxwell, S. E., & Cole, D. A. (2007). Bias in cross-sectional analyses of longitudinal mediation. *Psychological Methods*, 12, 23-44.
- McArdle, J. J. (1988). Dynamic but structural equation modeling of repeated measures data. In J. R. Nesselroade, & R. B. Cattell (Eds.), *Handbook of multivariate experimental psychology* (2nd ed., pp. 561-614). New York: Plenum Press.

- McArdle, J. J. (2009). Latent variable modeling of differences and changes with longitudinal data. *Annual Review of Psychology*, 60, 577-605.
- Mittelman, M. S., Haley, W. E., Clay, O. J., & Roth, D. L. (2006). Improving caregiver well-being delays nursing home placement of patients with Alzheimer disease. *Neurology*, 67, 1592-1599.
- Muller, D., Judd, C. M., & Yzerbyt, V. Y. (2005). When moderation is mediated and mediation is moderated. *Journal of Personality and Social Psychology*, 89, 852-863.
- Muthén, L. K., & Muthén, B. O. (2002). How to use a Monte Carlo study to decide on sample size and determine power. *Structural Equation Modeling*, 9, 599-620.
- Paxton, P., Curran, P. J., Bollen, K. A., Kirby, J., & Chen F. (2001). Monte Carlo experiments: Design and implementation. *Structural Equation Modeling*, 8, 287-312.
- Plotnikoff, R. C., Pickering, M. A., Flaman, L. M., & Spence, J. C. (2010). The role of self-efficacy on the relationship between the workplace environment and physical activity: A longitudinal mediation analysis. *Health Education & Behavior*, 37, 170-185.
- 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., & Hayes, A. F. (2004). SPSS and SAS procedures for estimating indirect effects in simple mediation models. *Behavior Research Methods, Instruments & Computers*, 36, 717-731.
- Prentice, R. L. (1989). Surrogate endpoints in clinical trials: Definition and operational criteria. *Statistics in Medicine*, 8, 431-440.
- Shrout, P. E., & Bolger, N. (2002). Mediation in experimental and nonexperimental studies: New procedures and recommendations. *Psychological Methods*, 7, 422-445.
- Raykov, T., Brennan, M., Reinhardt, J. P., & Horowitz, A. (2008). Comparison of mediated effects: A correlation structure modeling approach. *Structural Equation Modeling*, 15, 603-626.
- Raykov, T. (1998). Satisfying a simplex structure is simpler than it should be: A latent curve analysis revisit. *Multivariate Behavior Research*, 33, 343-363.
- Raykov, T. (2000). On sensitivity of structural equation modeling to latent relation misspecification. *Structural Equation Modeling*, 7, 596-607.
- Rogosa, D., & Willett, J. B. (1985). Satisfying a simplex structure is simpler than it should be. *Journal of Educational Statistics*, 10, 99-107.
- Roth, D. L., Mittelman, M. S., Clay, O. J., Madan, A., & Haley, W. E. (2005). Changes in social support as mediators of the impact of a psychological intervention for spouse caregivers of persons with Alzheimer's disease. *Psychology and Aging*, 20, 634-644.
- Roth, D. L., & MacKinnon, D. P. (2012). Mediation analysis with longitudinal data. In J. T. Newsom, R. N. Jones, & S. M. Hofer (Eds.), *Longitudinal data analysis: A practical guide for researchers in aging, health, and social sciences* (pp. 181-216). New York: Routledge.

- Rothman, K. J., Greenland, S., & Lash, T.L. (2008). Validity in epidemiologic studies. In K. J. Rothman, S. Greenland, & T. L. Lash TL (Eds.), *Modern epidemiology* (3rd ed, pp. 128-147). Philadelphia: Lippincott Williams & Wilkins.
- Sayer, A. G., & Cumsille, P. E. (2001). Second-order latent growth models. In L. M. Collins & A. G. Sayer (Eds.), *New methods for the analysis of change* (pp. 179-200). Washington, DC: American Psychological Association.
- Selig, J. P., & Preacher, K. J. (2009). Mediation models for longitudinal data in developmental research. *Research in Human Development*, 6, 144-164.
- Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence*. Oxford: Oxford University Press.
- Skrondal, A., & Rabe-Hesketh, S. (2004). *Generalized latent variable modeling: Multilevel, longitudinal, and structural equation models*. Boca Raton, FL: Chapman & Hall/CRC.
- Sobel, M. E. (1982). Asymptotic confidence intervals for indirect effects in structural equation models. *Sociological Methodology*, 13, 290-312.
- Spearman, C. (1904). The proof and measurement of association between two things. *American Journal of Psychology*, 15, 72-101.
- Stephenson, M. T., & Holbert, R. L. (2003). A Monte Carlo simulation of observable versus latent variable structural equation modeling techniques. *Communication Research*, 30, 332-354.
- Taylor, A. B., MacKinnon, D. P., Tein J. Y. (2008). Tests of the three-path mediated effect. *Organizational Research Methods*, 11, 241-269.
- Tein, J. Y., & MacKinnon, D. P. (2003). Estimating mediated effects with survival data. In H. Yanai, A. O. Rikkyo, K. Shigemasu, Y. Kano, J. J. Meulman (Eds.), *New developments in psychometrics: Proceedings of the International Meeting of the Psychometric Society* (pp. 405-412). Tokyo: Springer-Verlag.
- Thiébaud, R., Jacqmin-Gadda, H., Chêne, G., Leport, C., & Commenges, D. (2002). Bivariate linear mixed models using SAS proc MIXED. *Computer Methods and Programs in Biomedicine*, 69, 249-256.
- Thoemmes, F., MacKinnon, D. P., & Reiser, M. R. (2010). Power analysis for complex mediational designs using Monte Carlo methods. *Structural Equation Modeling*, 17, 510-534.
- Vittinghoff, E., Sen, S., & McCulloch, C. E. (2009). Sample size calculations for evaluating mediation. *Statistics in Medicine*, 28, 541-557.
- von Soest, T., & Hagtvet, K. A. (2011). Mediation analysis in a latent growth curve modeling framework. *Structural Equation Modeling*, 18, 289-314.
- Wilson, I. B., & Cleary, P. D. (1995). Linking clinical variables with health-related quality of life: A conceptual model of patient outcomes. *JAMA*, 272, 59-65.

APPENDIX

UAB INSTITUTIONAL REVIEW BOARD (IRB) APPROVAL FORMS FOR FIRST PREPRINT



Institutional Review Board for Human Use

Form 4: IRB Approval Form
Identification and Certification of Research
Projects Involving Human Subjects

UAB's Institutional Review Boards for Human Use (IRBs) have an approved Federalwide Assurance with the Office for Human Research Protections (OHRP). The Assurance number is FWA00005960 and it expires on October 26, 2010. The UAB IRBs are also in compliance with 21 CFR Parts 50 and 56 and ICH GCP Guidelines.

Principal Investigator: BENTLEY, JOHN P
Co-Investigator(s): BROWN, CYNTHIA J
Protocol Number: **X100810001**
Protocol Title: *Functional Status, Life-Space Mobility, and Quality of Life: Testing Mediation with Longitudinal Data*

The IRB reviewed and approved the above named project on 8-10-10. The review was conducted in accordance with UAB's Assurance of Compliance approved by the Department of Health and Human Services. This Project will be subject to Annual continuing review as provided in that Assurance.

This project received EXPEDITED review.

IRB Approval Date: 8-10-10

Date IRB Approval Issued: 8-10-10

Marilyn Doss, M.A.
Vice Chair of the Institutional Review
Board for Human Use (IRB)

Investigators please note:

The IRB approved consent form used in the study must contain the IRB approval date and expiration date.

IRB approval is given for one year unless otherwise noted. For projects subject to annual review research activities may not continue past the one year anniversary of the IRB approval date.

Any modifications in the study methodology, protocol and/or consent form must be submitted for review and approval to the IRB prior to implementation.

Adverse Events and/or unanticipated risks to subjects or others at UAB or other participating institutions must be reported promptly to the IRB.

470 Administration Building
701 20th Street South
205.934.3789
Fax 205.934.1301
irb@uab.edu

The University of
Alabama at Birmingham
Mailing Address:
AB 470
1530 3RD AVE S
BIRMINGHAM AL 35294-0104

Form 4: IRB Approval Form
Identification and Certification of Research
Projects Involving Human Subjects

UAB's Institutional Review Boards for Human Use (IRBs) have an approved Federalwide Assurance with the Office for Human Research Protections (OHRP). The Assurance number is FWA00005960 and it expires on September 29, 2013. The UAB IRBs are also in compliance with 21 CFR Parts 50 and 56.


Principal Investigator: BENTLEY, JOHN P
Co-Investigator(s): BROWN, CYNTHIA J
Protocol Number: **X100810001**
Protocol Title: *Functional Status, Life-Space Mobility, and Quality of Life: Testing Mediation with Longitudinal Data*

The IRB reviewed and approved the above named project on 8-16-11. The review was conducted in accordance with UAB's Assurance of Compliance approved by the Department of Health and Human Services. This Project will be subject to Annual continuing review as provided in that Assurance.

This project received EXPEDITED review.

IRB Approval Date: 8-16-11

Date IRB Approval Issued: 8-16-11


Marilyn Doss, M.A.
Vice Chair of the Institutional Review
Board for Human Use (IRB)

Investigators please note:

The IRB approved consent form used in the study must contain the IRB approval date and expiration date.

IRB approval is given for one year unless otherwise noted. For projects subject to annual review research activities may not continue past the one year anniversary of the IRB approval date.

Any modifications in the study methodology, protocol and/or consent form must be submitted for review and approval to the IRB prior to implementation.

Adverse Events and/or unanticipated risks to subjects or others at UAB or other participating institutions must be reported promptly to the IRB.

470 Administration Building
701 20th Street South
205.934.3789
Fax 205.934.1301
irb@uab.edu

The University of
Alabama at Birmingham
Mailing Address:
AB 470
1530 3RD AVE S
BIRMINGHAM AL 35294-0104