King's Research Portal

[Link to publication record in King's Research Portal](#)

Date of article: 08/05/2017

Tutorial: The Practical Application of Longitudinal Structural Equation Mediation Models in Clinical Trials

Kimberley A. Goldsmith

King's College London

David P. MacKinnon

Arizona State University

Trudie Chalder

King's College London

Peter D. White

Queen Mary University

Michael Sharpe

University of Oxford

Andrew Pickles

King's College London

Author Note.

Kimberley A. Goldsmith, Biostatistics & Health Informatics Department, King's College London; David P. MacKinnon, Department of Psychology, Arizona State University; Trudie Chalder, Academic Department of Psychological Medicine, King's College London; Peter D. White, Wolfson Institute of Preventive Medicine, Queen Mary University; Michael Sharpe, Department of Psychiatry, University of Oxford; Andrew Pickles, Biostatistics & Health Informatics Department, King's College London.

Correspondence concerning this article should be addressed to Kimberley Goldsmith, Room E2.03, Biostatistics & Health Informatics Department, Institute of Psychiatry, Psychology & Neuroscience, King's College London, DeCrespigny Park, London, UK, SE5 8AF.

E-mail: kimberley.goldsmith@kcl.ac.uk

Abstract

The study of mediation of treatment effects, or how treatments work, is important to understanding and improving psychological and behavioral treatments, but applications have often focused on mediators and outcomes measured at a single time point. Such cross-sectional analyses do not respect the implied temporal ordering that mediation suggests. Clinical trials of treatments often provide repeated measures of outcomes and, increasingly, of mediators as well. A trial with repeated measurements allows for the application of various types of longitudinal structural equation mediation models. These provide for flexibility in modeling, including the incorporation of some types of measurement error and unmeasured confounding that can strengthen the robustness of findings. The usual approach is to identify the most theoretically plausible model and apply that model. In the absence of clear theory, we put forward the option of fitting a few theoretically plausible models, providing a type of sensitivity analysis for the mediation hypothesis. In this tutorial, we outline how to fit several longitudinal mediation models. This will allow readers to learn about one type of model that is of interest, or about several alternative models so that they can take this sensitivity approach. We use the "**P**acing, Graded **A**ctivity, and **C**ognitive Behavioral Therapy: A Randomized **E**valuation" (PACE) trial of rehabilitative treatments for chronic fatigue syndrome (ISRCTN 54285094) as a motivating example and describe how to fit and interpret various longitudinal mediation models using simulated data similar to those in the PACE trial. The simulated dataset and Mplus code and output are provided.

Keywords: mediation, longitudinal mediation models, structural equation models, measurement error, clinical trials, chronic fatigue syndrome

Clinical trials of psychological and behavioral treatments are large and expensive experiments, making it essential that we use them to extract as much knowledge as possible. Thus it is important to learn not only whether a treatment works, but also how it works, and if it does not work where in the assumed therapeutic pathway it fails. The question of how treatments work can be addressed using mediation analysis (Baron & Kenny, 1986; Judd & Kenny, 1981; MacKinnon & Dwyer, 1993). This method allows the quantification of the effect of treatment on mediator and outcome variables, as well as the effect of treatment on outcome through one or more mediators.

Many studies assess mediation using single measurements of the putative mediator and outcome variables taken at the same time, which cannot in general provide causal mediation estimates due to bias (Maxwell & Cole, 2007; Maxwell, Cole, & Mitchell, 2011). When we have a mechanistic mediation hypothesis, it indicates that we think the treatment causes a change in the intermediate we have targeted, which in turn causes a change in an outcome of interest. So a mediation hypothesis implies a longitudinal model where treatment, mediator and outcome variables should be measured at three separated and ordered time points. In a clinical trial setting, we would apply the treatment at baseline, measure the mediator at an initial post-randomization time point, and the outcome at a later post-randomization time point. We also need to measure the variables on a timeline such that we capture sequential change, the mediator being measured early enough following intervention before change starts to take place in the outcome. For an extensive discussion of design and timing considerations in longitudinal studies, see the work of Gollob and Reichardt, and Cole and Maxwell (Cole & Maxwell, 2003; Gollob & Reichardt, 1987, 1991; Maxwell & Cole, 2007; Maxwell et al., 2011).

The availability of repeated measures in clinical trials allows the application of longitudinal structural equation models (SEM) of mediation to more fully model the mediator

and outcome processes together (MacKinnon, 2008). Three major defining features of structural equation models (SEM) are: (a) the ability to handle repeated measures of variables, (b) simultaneous estimation of multiple equations allowing the study of direct and indirect effects, and (c) the incorporation of measurement error using latent variables (Bollen & Noble, 2011; Kline, 2011). Within this framework, trajectories of mediators and outcomes can be modelled over time, and models may be able to account in part for potential sources of bias, such as some types of measurement error, and measured and unmeasured confounding (Goldsmith, Chalder, White, Sharpe, & Pickles, 2016).

The usual approach when applying SEM in general, and these longitudinal mediation models in particular, is to choose the most theoretically plausible model to apply and evaluate. However, ensuring the correct specification of these complex models can be challenging, and the models often make untestable assumptions. Another option is therefore to use theory as much as possible, in combination with the fitting of models from more than one theoretically plausible longitudinal model class. These models can then be used to explore specifications and assumptions by studying model fit and parameters of interest. While choosing a model based on theory is generally preferred, we suggest fitting a few competing plausible alternatives is also useful in allowing for a sensitivity analysis of mediation estimates that span an often ignored domain of uncertainty – selecting the correct model structure. As always, a cautious reasoned case should be made before giving selective prominence to one post-hoc preferred set of results, and full transparent reporting is required. This manuscript uses a tutorial format to describe several types of longitudinal models for mediation, allowing either approach to be taken.

We describe four alternative models for longitudinal mediation: (a) simplex with lagged mediation paths, (b) simplex with contemporaneous mediation paths, (c) latent growth, and (d) a modified latent change model with a less restrictive parameterization.

These models have been described previously in the literature (Cheong, MacKinnon, & Khoo, 2003; Dunn, Everitt, & Pickles, 1993; Jöreskog, 1970; Krull, Cheong, Fritz, & MacKinnon, 2016; Marsh, 1993; McArdle, 2009; Neale & Cardon, 1992; Raykov, Marcoulides, & Boyd, 2003; Selig & Preacher, 2009; Steyer, Eid, & Schwenkmezger, 1997; Willett & Sayer, 1994), but we know of no previous publication applying, comparing and contrasting these models in relation to mediation, nor providing materials in a tutorial format.

The original motivation for this work was the "Pacing, Graded Activity, and Cognitive Behavioral Therapy: A Randomized Evaluation (PACE) trial" (White et al., 2011). The study of mediation was incorporated in the design, with mediators and outcomes measured at multiple time points, and we have created a simulated dataset based on this trial. Our aim is to facilitate the application of these models to researchers' own data, so we have provided Mplus software code, annotated output, and information on using Mplus to estimate mediation effects. We used the Mplus program for model fitting because it is in common use, can handle missing data using maximum likelihood and straightforwardly provides indirect effects and their associated standard errors/confidence intervals (CI). Users of other software programs should be readily able to translate the provided Mplus commands into their chosen software.

Purpose of Tutorial and Intended Audience

The purpose of the tutorial is to guide individuals who have some knowledge and experience of fitting single and/or cross-sectional mediator models, and who are interested in learning how to fit more complex longitudinal models for mediation. This could include PhD students, postdocs, and more advanced researchers familiar with SEM but not specifically with fitting longitudinal mediation models. The tutorial could also provide longitudinal SEM for mediation teaching material for instructors.

We focus on a common and relatively straightforward situation in which to assess causal mediation – a randomized clinical trial – however, the methods could be applied to quasi-experimental designs or evaluation of an exposure in an observational study, though the latter would likely require more cautious interpretation. As well as knowledge and some experience of fitting SEM, we will assume an understanding of Sewell Wright's path tracing rules (Wright, 1920a, 1920b). For those less familiar with mediation and SEM for mediation, we provide further background and references in the Supplemental Material.

Outline of the Tutorial

We begin by introducing a common notation for mediation, model assumptions, conceptualization of the mediation pathways, and the use of full information maximum likelihood estimation. We then outline the motivating example and describe the simulated dataset. We follow this with some important considerations for the longitudinal mediation models, including lagging of paths, allowing for confounding, further assumptions specific to these models, and a description of the mediated effects. We then consider each longitudinal mediation model in turn, explaining the calculation of mediated effects specific to that model, and their interpretation. Although in practice we would assess goodness-of-fit and choose among models before interpreting effects, we delay this until the end of the tutorial when all models of interest have been presented. Detailed practical guidance and the Mplus materials can be found in the Supplemental Material.

Notation and General Model Assumptions

We are focusing on the estimation of mediated effects in a clinical trial, where we have R as a dummy variable coding treatment versus control, M as a mediator, and Y as an outcome. The treatment R will have been designed to target an intermediate variable, M. We will refer to the path estimating the relationship between R and M as the *a* path, as is

generally done in the mediation literature.   We then hypothesize an effect of M on Y, and will refer to this as the *b* path. We generally assume that there could also remain a direct effect of R on Y; this is referred to as the *c'* path. We assume familiarity with a simple mediation models such as this, however, we provide further background and a diagram of such a model in the Supplemental Material (Figure S1). The Figures in the manuscript provide the necessary extensions to the notation in the case of longitudinal models where we have several M, Y, and *a*, *b* and *c'* paths.

In order to avoid bias in estimating a mediated effect, several other assumptions must hold. We assume (a) reliably and validly measured variables, (b) linear relationships between variables, including no R x M interaction on Y, and (c) no confounding of the M – Y relationship by post-randomization variables (De Stavola, Daniel, Ploubidis, & Micali, 2015; Robins & Greenland, 1992; VanderWeele, 2009; VanderWeele & Vansteelandt, 2009; VanderWeele, Vansteelandt, & Robins, 2014). In simple mediation models, the assumption of no unmeasured confounding of the mediator – outcome relationship must also be made, which could be a substantial source of bias (Emsley, Dunn, & White, 2010; Imai, Keele, & Tingley, 2010; Judd & Kenny, 1981; MacKinnon, 2008; MacKinnon & Pirlott, 2015; VanderWeele & Vansteelandt, 2009). Some unmeasured confounding could be allowed for via covarying M and Y. We cannot allow for such covariation in a simple mediation model because the model would not be identified, but as we explain later, we can allow for this sort of relationship in longitudinal mediation models (Goldsmith et al., 2016). In an observational study where the independent variable might be an exposure X, any of the relationships between X, M, and Y could be subject to unmeasured confounding (MacKinnon & Pirlott, 2015). We will focus exclusively in this tutorial on the case where R is a randomized treatment, the mediator and outcome are continuous measures, and we make assumptions (a)

through (c) stated above; readers are referred to the literature mentioned and the wider causal inference literature in cases where these assumptions may be not satisfied.

A particular situation where we may be more likely to have an observed X rather than randomized R is the program evaluation field, where interventions often cannot be randomized.   There is a useful paradigm from program theory and intervention evaluation for studying mediation both in trials and observational studies. This describes the *a* path as the "action theory", where we are taking action on a mediator using a treatment, intervention or program, and the *b* path as the "conceptual theory", where there is a "known" relationship between the mediator and outcome that we are trying to affect by changing the mediator (Chen, 1990). This paradigm facilitates the study of mediation whether causal or non-causal effects are studied (MacKinnon, 2008; MacKinnon, Lockwood, Hoffman, West, & Sheets, 2002; MacKinnon & Pirlott, 2015; Pek & Hoyle, 2016).

Although the simulated data are complete, missing data will be an issue in real datasets. We would generally assume data are missing at random and recommend using the full information maximum likelihood algorithm (FIML), as is available in Mplus, to fit models (Enders & Bandalos, 2001; Okleshen Peters & Enders, 2002). Using maximum likelihood and missing at random assumes that all variables predicting missing data are included in the model, so readers should ensure this. The missing at random assumption is likely more plausible in the longitudinal models demonstrated here because they incorporate repeated measurements, and earlier measures of the mediator/outcome likely predict later missing data. Data could of course be missing not at random; in this case appropriate sensitivity analyses and/or multiple imputation should be explored (Resseguier, Giorgi, & Paoletti, 2011; Zhang, Wang, & Tong, 2015).

Motivating Study – The PACE Trial

The Pacing, Graded Activity, and Cognitive Behavioral Therapy: A Randomized Evaluation (PACE) trial studied different rehabilitative therapies for chronic fatigue syndrome (CFS) (White et al., 2011). The PACE trial had four treatment arms. One was a control treatment, specialist medical care (SMC), which all trial participants received, with the other arms delivering cognitive behavioral therapy (CBT), graded exercise therapy (GET), and adaptive pacing therapy (APT) in addition to SMC. The study reported that both CBT and GET were more effective than APT and SMC in improving the two primary outcomes, fatigue and physical function (White et al., 2011). Interested readers are referred to protocol and primary analysis papers (White et al., 2011; White, Sharpe, Chalder, DeCesare, & Walwyn, 2007).

We were interested in mediation of the effects of the PACE treatments, and found evidence for mediation of treatment effects using simple mediator models (Chalder, Goldsmith, White, Sharpe, & Pickles, 2015). However, the PACE trial also provided an outstanding opportunity to apply longitudinal mediation models. Firstly, multiple treatment arms versus the usual two arm design of the trial allowed for wider exploration of mediation effects and assumptions. Secondly, both mediators and outcomes were measured at four time points – baseline (week 0), mid-treatment (12 weeks post-randomization), post-treatment (24 weeks post-randomization), and follow-up (52 weeks post-randomization). Such repeated measures over time allow more complex joint longitudinal modeling of the mediator and outcome processes (Cheong et al., 2003; Cole & Maxwell, 2003; Dunn et al., 1993; Goldsmith et al., 2016; MacKinnon, 2008; Marsh, 1993; McArdle, 2009; Pickles et al., 2015).

Simulated Data Provided for the Tutorial

Data were simulated using Mplus, with parameters from a model fitted to the PACE data that were modified before being used in the simulation algorithm (Supplemental Material, Figure S2). The data were simulated based on PACE, but are not actual PACE data. The dataset contains the binary 0, 1 coded variables r1, r2, r3, and r4, which represent four treatment groups; four mediator measurements, m0 at baseline and three post-randomization time points, m1, m2, and m3; and corresponding outcome measures, y0, y1, y2, and y3. The r1 – r4 variables correspond to the CBT (R1), APT (R2), and GET (R3), and SMC control treatment (R4). R1 and R3 are therefore active treatments, with R2 as an example of a treatment that differs little from the control. The simulated mediator data was based on a measure of fear avoidance (Knudsen, Henderson, Harvey, & Chalder, 2011; Moss-Morris & Chalder, 2003; Skerrett & Moss-Morris, 2006), which is scored as higher is worse. The simulated outcome data was based on the physical functioning outcome (Buchwald, Pearlman, Umali, Schmaling, & Katon, 1996; McHorney, Ware, & Raczek, 1993), which is scored as higher is better. These simulated data and further detail on the simulation have been provided in the Supplemental Material. We have also provided a plot of the observed means over time in the PACE data in Figure S3 in the Supplemental Material to aid later discussion of trajectories of change.

The data used to generate the parameters used for simulation were standardized to baseline for each of the mediator and outcome. We will leave whether and how to standardize to the discretion of the reader as there is some disagreement on this subject in the literature (Baguley, 2009; Cheung, 2009). Standardizing gives effects that are arguably in more interpretable units of measurement, i.e. standard deviation (SD) units, as compared to the original scales of the measures (Cheung, 2009; MacKinnon, 2008). Standardization in mediation helps with the use of different scales for mediator and outcome, and potentially

allows for cross-measure comparisons. Using baseline values to standardize provides

measurement units that are retained across time, scaled independently of the treatment receipt

effects, and indirect effects expressed in units of baseline SD of the outcome (i.e. the SD of

the mediator cancels out when calculating the indirect a x b effect). Standardization may be

particularly important for the indirect effect, which as a product of two estimates may be

challenging to interpret if calculated using the original variable scales (Cheung, 2009;

MacKinnon, 2008; Preacher & Kelley, 2011).

Longitudinal Structural Equation Models for Mediation

We will now present some important general longitudinal mediation model

considerations, followed by four SEM that could be used to model the mediator and outcome

processes. The models are simplex models with (a) lagged, and (b) contemporaneous b path

effects, (c) a latent change model, and (d) a modified latent change model. We recommend

all readers go through the simplex with lagged b paths model section, where we have

provided a more detailed explanation of the effects of interest and the meaning of the

mediation findings. The descriptions provided for the other models are briefer. To model

treatment, we are entering dummy variables R1 – R3 with R4 included in the dataset, i.e. this

provides contrast effects for each of the treatments as compared to the control. For the

purposes of the tutorial we focus on overall indirect and direct effects at the third post-

randomization time point. The models are shown in Figures 1 through 4. The figures include

products of coefficients expressions for indirect and direct effects for the third time point, and

estimates of important paths and mediated effects with their 95% CI. The numbers in round

brackets in the figures are standard errors, with CI in square brackets. The effects for the

first, second, and third post-randomization time points are all calculated in the Mplus code in

the Supplemental Material, with results for the first two time points shown in Supplemental

Tables S1 through S5.

Lagged Versus Contemporaneous *b* Paths

Lagged paths are those between an earlier measure of the mediator and a later

measure of the outcome, for example, the b_L paths in the simplex model in Figure 1.

Although models with lagged *b* path effects make more chronological sense theoretically, in

our and others' experience, contemporaneous *b* path effects between mediator and outcome at

the same time point, i.e. the b_C paths in Figure 2, may give better fitting models (Goldsmith et

al., 2016; Wang, Zhang, & Estabrook, 2009). This may especially be the case where there is

a substantial delay before the first post-randomization measurement of the mediator. Most

clinical trials tend to have this sort of delay, as most take their earliest measure of the

mediator either at the end of the end of the most intensive phase of therapy, or at the end of

therapy altogether. It is likely that most of the change in the mediator, and potentially in the

outcome, has already occurred by that time, leading to stronger relationships between

measures taken at the same time point. We consider both a simplex model with lagged *b*

paths and a simplex model with contemporaneous *b* paths.

Allowing For Confounding in Models – Mediator/Outcome Residual Covariance

While the effect estimates for the *a* path are readily interpreted as causal, because of

randomized assignment to treatment, the post-randomization *b* path estimates are much more

vulnerable to confounder and measurement error bias (regression attenuation). To avoid the

cross-sectional association of mediator and outcome at baseline contaminating the *b* path

estimates, it is now recognized that all models should adjust for baseline mediator and

outcome variables (Dunn, Emsley, Liu, & Landau, 2013; Pickles et al., 2015). However

other, possibly unobserved, variables may give rise to additional confounding covariance between mediator and outcome (Goldsmith et al., 2016). Some longitudinal mediation models will remain identified when we allow for covariance between mediator and outcome by means of correlated residual terms or by a shared latent variable. There are then different options for where such additional covariance can be allowed (a) between the mediator at one time point and the outcome at the following time point (lagged), (b) between mediator and outcome measured at the same time point (contemporaneous), or (c) both, if such models were identified. We have allowed for contemporaneous covariance paths in the models considered here because previous findings were more in support of covariances between mediator and outcome at the same time point (Goldsmith et al., 2016). This could be due to the same individuals reporting both mediator and outcome, generating correlations in the occasion-specific residuals.

Further Longitudinal Model Assumptions and Considerations

Our previous work with longitudinal simplex mediation models in the PACE data suggested some reasonable simplifications which we have applied here (Goldsmith et al., 2016), (a) setting the *b* paths equal over time, (b) setting the *b* paths equal across treatments, i.e. assuming no R x M interactions on Y, (c) setting mediator and outcome residual variances to be equal over time except in the latent growth model, (d) setting the treatment to the third measure of mediator and outcome paths equal to zero, since treatment had finished before the second set of measurements were taken, and (e) setting the mediator – outcome residual covariances to be equal over time. Readers will need to assess the plausibility of similar assumptions in their own studies where possible.

Setting the *b* paths to be equal is parsimonious and can offer more precise estimation. Theoretical support for this can be offered by returning to the program evaluation paradigm

of the *a* and *b* paths as action and conceptual theories, with the latter constituting a known relationship we expect to be affected by changing the mediator (Chen, 1990). This suggests the conceptual theory relationship between mediator and outcome should be reasonably consistent over time and across treatments, i.e. no matter how the mediator is changed, it has a similar effect on the outcome. We would generally look to the literature and/or a number of our own studies for evidence of this mediator – outcome relationship. We found evidence for the strong but appealing assumption of common *b* paths across treatments in previous work (Chalder et al., 2015), i.e. no R x M interactions on Y. Assuming common *b* paths across treatments and/or time could lead to bias if unjustified, so as mentioned, it is important to assess the plausibility of such assumptions.

Indirect (Mediated) and Direct Effects in Longitudinal Mediation Models

Longitudinal mediation models with repeated measures of both mediator and outcome allow for many different indirect effects. Cole and Maxwell (2003) expanded on work by Gollob and Reichardt (1991), specifically defining effects for longitudinal mediation models, and we will define similar effects for the models described in this tutorial. In models with repeated mediator and outcome measurements, indirect/mediated effects include all paths that go from treatment to outcome through any measure of the mediator. Direct effects are paths that go from treatment to outcome that can pass through several measures of the outcome, but don't pass through any measures of the mediator (Cole & Maxwell, 2003; Gollob & Reichardt, 1991).

Cole and Maxwell also further classify effects as either time-specific effects that are estimated and calculated for a specific time point, and overall effects up to and including the final follow-up measure, with the overall effects generally being of most interest (Cole & Maxwell, 2003). So for our example, the overall indirect effect includes all time-specific indirect effects for the third post-randomization time point. Another way to describe this is

the overall indirect effect includes all pathways from randomized treatment to the measurement of the outcome at the third post-randomization time point that pass through any measure of the mediator. These are the effects we focus on.

Since indirect effects are products of coefficients for which normal theory precision estimates perform poorly, it is recommended that CI for the indirect effects be obtained by percentile bootstrap (Fritz, Taylor, & MacKinnon, 2012; MacKinnon, Lockwood, & Williams, 2004). We used 1000 bootstrap replications to get CI for these effects (Efron & Tibshirani, 1993).

Simplex Models for Mediation

Simplex models have each observed value of the mediator and outcome as single indicators for a latent 'true score' factor, as shown in Figure 1. Each factor includes an occasion specific residual term whose variation will include some forms of measurement error (Dunn et al., 1993; Jöreskog, 1970, 1979; Marsh, 1993). The models partition the true score and residual from one another using a decomposition of covariances (Kline, 2011). These models then usually postulate a first-order autoregressive structure among the true scores where a variable is a function of that variable at the previous time point, resulting in the correlation between measurements decreasing the further apart they are in time. For longitudinal mediation, the simplex structure is fitted to each of the mediator measures and outcome measures and then the processes are joined through *b* paths between the latent variables (Figures 1 and 2). These *b* paths can be lagged (Figure 1), contemporaneous (Figure 2), or both. The *a* paths in these models are those between the treatment group variable(s) (the R dummy variables) and the latent mediator at each time point.

Simplex models for a single process require at least three measurements, and with only a single indicator for each latent variable, are generally not identified without further

constraints (Dunn et al., 1993; Jöreskog, 1970; Marsh, 1993). A straightforward and

plausible option suggested in the literature and which we use here is to assume factor

loadings are all equal to 1 with constant residual variances over time (Dunn et al., 1993;

Jöreskog, 1970; Marsh, 1993). Another option not applied here is to use an estimate of

reliability should one be available (Bagozzi & Heatherton, 1994; Bollen, 1989; MacKinnon,

2008; Stephenson & Holbert, 2003). We note that the measurements in the PACE trial were

unequally spaced, but we have used models assuming equal spacing for simplicity. Readers

could consider simplex models where the "missing" equally spaced time point is modeled as

a latent variable, (Dunn et al., 1993), or the parameter in question as some function of time,

both of which involve further assumptions. For example in Figure 1, the mediator

autoregressive paths could have been modeled as m, m, 2m to reflect the time spacing. It

might be better in such situations to use the latent growth or latent change models described

later, where unequal spacing of measurements can be more explicitly modeled.

Indirect (mediated) and direct effects in simplex mediation models

To calculate indirect effects for a given treatment group we need to choose the

appropriate a paths. In most of the models shown here, the indirect effects of R1 go via the

a_{11} and a_{12} paths. For R3, the effects go via a_{31} and a_{32}. One time-specific indirect effect at

the third post-randomization time point for the R1 treatment group in the simplex lagged

model is R1 -> FM_1 -> FM_2 -> FY_3 -> Y_3, or in following path tracing rules to estimate the

covariance, a_{11} x m_2 x b_L x 1 (Figure 1). These and the rest of the paths for the indirect and

direct effects are shown at the bottom of Figure 1. The corresponding effect for R3 is a_{31} x

m_2 x b_L x 1. Note that in this case we are calculating effects on the scale of measurement of

the standardized observed variables and so include the paths between the latent and observed

variables/factor loadings in this and other indirect and direct effect calculations (Sobel, 1986).

This can be seen in the last path of the above example, FY_3 -> Y_3. Readers may choose to calculate effects at the latent variable level and so would omit factor loadings from the calculations. Many of the factor loadings are = 1 and so make no difference to the calculation, but where this is the case we include these for completeness. To get the overall indirect effect for the third post-randomization time point for treatment R1, we calculate all of the time-specific indirect effects, i.e. the products of the parameters for each of the three paths between R1 and Y_3 that pass through a measure of the mediator (Figure 1), and then sum all of these products.

The residual direct effects of treatment are those that follow paths that do not pass through any measure of the mediator. An example of a direct effect for the R1 treatment at the third post-randomization time point is R1 -> FY_1 -> FY_2 -> FY_3 -> Y_3, or c'_{11} x y_2 x y_3 x 1 (Figure 1). As in the case of the indirect effects, the overall direct effect for this time point is composed of the sum of all the time-specific direct effects. Effects for the earlier time points are calculated in a similar way. We can calculate total effects at each time point as the total time-specific indirect effect plus the total time-specific direct effect. Note if we hadn't constrained the b paths to be the same for each post-randomization time point, we would have to substitute the appropriate b path in these calculations. Please see the Supplemental Material for the calculation of these effects using the Mplus program ("Fitting Longitudinal Mediation Models Using Mplus" section).

Results of fitting a simplex model with lagged mediation paths

When fitting a simplex model with lagged b paths to the simulated data, we see that two of the of the a_1 path estimates were statistically significant (Figure 1). The estimate for the R1 treatment group (a_{11} path) was -0.82, 95% CI [-0.63, -1.02], and for the R3 treatment group (a_{31} path) the estimate was -0.96 [-0.79, -1.17] (Figure 1, upper table). These effects

mean that R1 and R3 decreased (improved) the simulated avoidance of fearful situations

mediator by 0.83 and 0.96 baseline mediator SD units as compared to the control treatment

R4. The a_{21} path for R2 was not significant, suggesting this treatment group had no effect on

the mediator as compared to the control. This was true throughout, so we will focus on the

R1 and R3 treatments. The estimate of the common b_L path had a magnitude of -0.05 [-

0.001, -0.10], and so was borderline statistically significant (i.e. the lower limit of the CI only

just excluded zero). The interpretation of this b path is that each baseline SD unit increase in

the mediator led to a 0.05 baseline SD unit decrease (worsening) in the physical functioning

outcome. While the a paths show an effect of treatment on the mediator, the small magnitude

of the b_L path (Figure 1, upper table) suggests only a small effect of earlier measures of the

mediator on later measures of the outcome. The b_L path was significant, however, and

significant a and b paths suggest that the indirect/mediated effects will be significant. Given

the small magnitude of the b_L path we would expect these indirect effects to be relatively

small.

In fact, we do find the overall indirect effects for R1 and R3 for the third time point to

be relatively small in magnitude and the confidence intervals to barely exclude zero with the

effect for R1 0.08 95% CI [0.001, 0.16], and for R3 0.09 [0.002, 0.18] (Figure 1, lower table).

The overall direct effects for the third time point for R1 and R3 were 0.71 [0.46, 0.97], and

0.66 [0.39, 0.93] (Figure 1, lower table). Taken overall, the results suggest that the effects of

R1 and R3 were partially and weakly mediated (Baron & Kenny, 1986), i.e. only part of the

total effect was mediated, with small but significant indirect effects and large and more

significant residual direct effects. With full mediation of a total treatment effect, we would

expect non-significant residual direct effects essentially equal to zero.

We further note the lack of a mediated effect of the R2 treatment, which we had

simulated to be ineffective, i.e. to have no total effect of treatment on the outcome. As

alluded to earlier, we can use mediation analysis to help clarify why this treatment was ineffective. We can see that this treatment did not appreciably affect the mediator (non-significant a_{21} path, Figure 1). In substantive applications, this would provide important information that could be used to refine the treatment.

Readers will need to think through the relationships in their data so as to check that effect estimates make sense. Here we note that the estimates of the *a* and *b* paths take on negative values, while the total, indirect and direct effects are all positive (for example, see Figure 1). The outcome is a simulated version of a physical functioning variable with scoring higher = better. Hence the total effect of the effective treatments on the outcome was positive, i.e. they improved the outcome. In calculating the direct and indirect effects, we partition the total effect, so we would generally expect both of these effects to be positive unless we have inconsistent mediation (MacKinnon, 2008). Here we had negative *a* and *b* paths; when we multiply these together, we arrive at a positive indirect effect.

Results of fitting a simplex model with contemporaneous mediation paths

The simplex model with contemporaneous *b* paths is shown in Figure 2. The R1 and R3 a_1 path estimates were very similar to those estimated by the lagged model, with the estimate for R1 (a_{11} path) being -0.82, 95% CI [-0.63, -1.01], and for R3 (a_{31} path) -0.96 [-0.79, -1.17] (Figure 2, upper table). The estimate of the common b_C path had a magnitude of -0.08 [-0.04, -0.12]. These *a* and *b* paths can be interpreted as for the simplex lagged model, except that the b_C path quantifies the mediator – outcome relationship at the same time point.

The significance of the *a* and *b* paths estimates suggests there will be significant mediated effects, as in the case of the simplex lagged model, but the somewhat larger b_C path estimate leads to indirect effects of larger magnitude, for R1 being 0.19 95% CI [0.09, 0.28], and for R3 0.21 [0.11, 0.32] (Figure 2, lower table). The overall residual direct effects for the

third time point for R1 and R3 were significant, with estimates of 0.62 [0.37, 0.87], and 0.55

[0.28, 0.81] (Figure 2, lower table). These effects can also be interpreted similarly to those

from the simplex lagged model. The indirect effects are of larger magnitude here,

representing partial moderate mediation and giving stronger evidence of mediation of the

effects of R1 and R3.

Latent Growth Models for Mediation

Longitudinal data can also be modeled using latent growth models in the SEM

framework, where the heterogeneous trajectories of individuals are modeled with latent

random intercepts and slopes, as shown in Figure 3 (Cheong et al., 2003; Dunn et al., 1993;

MacKinnon, 2008; Muthen & Curran, 1997). Typically, the latent intercept variable loads on

each observed variable with a factor loading of one, whereas the latent slope variable is given

loadings to reflect a plausible trajectory. For example, to model four equally spaced

measures with a linear slope, the loadings could be [0, 1, 2, 3] (Cheong et al., 2003; Muthen

& Curran, 1997). A square root transformation of the linear loadings, for example [0, 1, 1.41,

1.73], can linearize a pattern over time where there is greater early change (Hedeker &

Gibbons, 1997). Other options include modeling a trajectory with early linear or step change

followed by a plateau, e.g. [0, 1, 1, 1], or with the benefit of repeated measurements, some

loadings may be estimated as free parameters (Cheong et al., 2003; Lockhart, MacKinnon, &

Ohlrich, 2011; MacKinnon, 2008), dependent upon the available degrees of freedom (i.e.

model identification). This or other approaches could be useful where there is unequal

spacing between measurements (Biesanz, Deeb-Sossa, Papadakis, Bollen, & Curran, 2004).

It has been suggested that the optimal trajectory should be determined separately for each of

the mediator and outcome processes before incorporating them together in a model, and we

refer readers to the literature for more information (Cheong et al., 2003; MacKinnon, 2008).

The *a* paths in the latent growth models are from the treatment variable(s) (R dummy variables) to the slope of the mediator, with the *b* path joining the slope of the mediator to the slope of the outcome (Figure 3) (Cheong et al., 2003; MacKinnon, 2008). The *b* path represents the relationship between the rate of change in the mediator process and the rate of change in the outcome process, each incorporating all repeated measures. In this sense, the temporal order criterion for mediation is not obviously met for the latent growth models, because the change in the mediator and outcome are both reduced to single slope estimates incorporating pre-treatment baseline measures.

In this tutorial, we have fitted a model with the factor loadings set as the square root of the time point, with the final loading parameter free/estimated in this model. This was done to address unevenly spaced time points, and the pattern of early change followed by a plateau as seen in the PACE trial data (see Supplemental Material, Figure S3).

Another implication of the linear latent growth models is that where the direction of the trajectory is specified, the models force an individual's predicted progression in time to maintain that same direction. This can be relaxed somewhat using some of the more flexible loadings specifications of change over time or slopes for non-linear effects of time (Cheong et al., 2003). Note that we do not need to assume equal residual variances over time in the latent growth models.

Indirect (mediated) and direct effects in latent growth mediation models

In the latent growth models there is one indirect and one direct effect for each treatment group at each time point. The indirect effect is *a* x *b* x the applicable factor loading (Cheong et al., 2003). For example, for a model with a linear slope, the indirect effect for R1 at the first post-randomization time point would be calculated as R1 -> SM -> SY -> Y_1 or a_1 x *b* x 1, the second post-randomization time point as a_1 x *b* x 2, and at the third post-

randomization time point as a_1 x b x 3. The indirect effects for R1 for the square root

parameterization with the final factor loading to be estimated are a_1 x b x 1 for the first post-

randomization time point, a_1 x b x 1.41 for the second post-randomization time point, and a_1

x b x 1.39 for the third post-randomization time point (Figure 3, lower table). The direct

effect is the path from treatment to the slope of the outcome (e.g. c'_1) multiplied by the factor

loading for the time point of interest, so for R1 for the third time point it is c'_1 x 3 in a linear

slope model, and c'_1 x 1.39 in the square root transformation model in Figure 3. It is useful

to calculate time specific effects as we have done for the other models. However, we note we

should interpret these effects recalling that the latent growth models summarize all the

measurements across time into a single slope parameter, and so don't maintain temporal M –

Y separation.

Results of fitting a latent growth model

        When fitting a latent growth mediation model, we saw that the estimates of the a

paths between each of the R1 and R3 treatment groups and slope of the mediator (a_1 and a_3)

were statistically significant, with magnitudes of -0.62 95% CI [-0.50, -0.76] for R1 and -0.70

for R3 [-0.57, -0.84] (Figure 3, upper table). In this case, the interpretation is the treatments

have decreased the slope of the mediator by 0.63 and 0.70 baseline mediator SD units. The

estimate of the b path -0.36 [-0.17, -0.74] was also significant, so we expect significant

indirect effects for both treatments. The b path can be interpreted as a one unit difference in

the slope of the mediator being associated with a -0.36 unit difference in the slope of the

outcome.

        The indirect effects for the third time point were substantial and significant, 0.31 for

R1 95% CI [0.14, 0.63], and 0.35 for R3 [0.16, 0.73] (Figure 3, lower table). The residual

direct effect for R1 was 0.46 for R1 [0.07, 0.75], with R3 0.35 for R3 [-0.08, 0.67]. These

results suggest a relatively higher level of mediation than the simplex models, to the extent

that complete mediation could be consistent with the results for R3 (i.e. the direct effect CI

does not rule out an effect equal to zero).

Latent Change Models for Mediation

Another class of longitudinal model that can be used to explore mediation is one that

allows change to differ over different time periods (Dunn et al., 1993; Lockhart et al., 2011;

Steyer et al., 1997). Similar models have been described by McArdle and colleagues

(McArdle, 2009). The models used here were parameterized as shown in Figure 4 and

Supplementary Material Figure S4, as random walk/Weiner models (Dunn et al., 1993). In

these models, the first latent variable loads on all four observed variables; the second latent

variable on the second, third, and fourth observations; the third latent variable on the third

and fourth observations; and the fourth latent variable just on the final observation, with all

factor loadings equal to one (Dunn et al., 1993; Steyer et al., 1997), as shown in the

straightforward latent change model shown in Figure S4 in the Supplemental Material. This

parameterization gives latent variables representing change scores between each time point

(Steyer et al., 1997). With four measurements of mediator and outcome, a modified latent

change model can also be fitted, where the loading on the observed variable at the same time

point is equal to one and the rest are estimated but constrained to be equal. These models

allow each new increment/decrement to the latent score to include a transient component that

does not necessarily persist beyond the next measurement. We focus on this model in the

tutorial.

In the latent change models, the *a* paths are from the treatment group variable(s) to the

latent change scores in the mediator. The *b* paths then join the latent change scores for the

mediator with the latent change scores for the outcome, in a similar fashion to the simplex

models. The *b* paths represent the relationship between change in the mediator and change in the outcome between two time points. These models have been fitted in the tutorial with contemporaneous *b* paths, as these were also found to be more plausible for the PACE data than lagged *b* paths in these change models (data not shown). The *b* paths therefore represent the relationship between mediator and outcome change over the same time period.

The latent variables in the latent change model are assumed to be independent, i.e. the change between each time point is modelled independently. At each time point a new latent variable contributes an increment/decrement to the latent score which then contributes undiminished to all later latent scores. Since the direction of change between each time point is independent, the latent score can increase over one time period and decrease over the next. Such a model could be suitable in general for trials of complex psychological therapies or interventions, where the greatest change in the outcome often occurs early. Another potential advantage of these models is they allow for different predictors of change at different time points.

Indirect (mediated) and direct effects in latent change mediation models

The indirect effects for the latent change models are derived in a similar fashion to those in the simplex models. The latent change models have a somewhat simpler path structure due to their assumption of independent latent change scores, however, these models have more complex factor loading paths, which come into the calculations for the modified latent change model.

For the modified latent change model shown in Figure 4, one time-specific indirect effect for the R1 treatment for the third post-randomization time point is R1 -> FM_1 -> FY_1 -> Y_3, or a_{11} x *b* x est_y. The second indirect effect for this time point is R1 -> FM_2 -> FY_2 -> Y_3, or a_{12} x *b* x est_y. In the fitted model in Figure 4 the factor loading is estimated as 1.02, so

these effects are a_{11} x b x 1.02 and a_{12} x b x 1.02. An example of a direct effect for R1 in the modified latent change model for the third time point is R1 -> FY_1 -> Y_3 or c'_{11} x est_y, with the other being R1 -> FY_2 -> Y_3 or c'_{12} x est_y.

Results of fitting latent change models

Estimates from the latent change model are shown in Figure S4 in the Supplemental Material, with effects from the modified latent change model shown in Figure 4 discussed here. We discuss these effects in terms of change for ease of expression, but readers should keep in mind that the modified model is more flexible and is not modeling absolute change, as is the case in the strictly latent change model. The estimates of the a_1 paths for R1 and R3 were statistically significant (a_{11} and a_{31} paths), and were -0.83 95% CI [-0.64, -1.01] for R1, and -0.95 [-0.78, -1.16] for R3 (Figure 4, upper table). The magnitude of the common b_C path was -0.32 [-0.18, -0.53]. The b path here can be interpreted as the effect of a one unit change in the mediator on a one unit change in the outcome.

The overall indirect effects at the third time point were 0.37 95% CI [0.21, 0.57] for R1, and 0.42 [0.26, 0.64] for R3 (Figure 4, lower table). The direct effects were 0.40 [0.09, 0.68] for R1, and 0.31 [-0.01, 0.59] for R3 (Figure 4, lower table). The indirect effects can be interpreted as the amount of the effect in baseline SD units of the total effect on the outcome that is mediated, for example for R3, 0.42 baseline SD units of the total effect on the outcome was mediated. The interpretation is similar to that for the latent growth model, i.e. the results are consistent with partial mediation of the effect of R1, and potentially complete mediation of the effect of R3, on the outcome.

Assessing model fit

As there is an ongoing controversy with regard to appropriate fit indices (Browne & Cudeck, 1992; Kline, 2011), we have pragmatically focused on a few. We use the Root Mean Square Error of Approximation and its associated 90% CI (RMSEA) (Steiger, 1990; Steiger & Lind, 1980), comparing to the generally accepted threshold of ≤ 0.05 for reasonable model fit (Hu & Bentler, 1999; Kline, 2011; Marsh & Hau, 1996). To make more formal comparisons, and to compare between non-nested models such as those applied here, we used the Bayesian and Akaike's Information Criteria (BIC and AIC) (Akaike, 1974; Schwarz, 1978), following the smaller is better criterion, and considering differences of 2 or greater to indicate meaningful differences between models (Kass & Raftery, 1995). We emphasize again that we are not advising readers to fit several models to pick the one showing the results most to the reader's liking, for example, the one with the largest mediated effect, but instead to use a combination of theory and model fit indices to put forward the most plausible models and results.

Given that the data were simulated from the modified latent change model, it is no surprise that fitting this model gave a low RMSEA approximately equal to zero 90% CI [<0.001, 0.020], implying a good fit. But what of the other models, which we know to be "wrong"? Could we have excluded them, and the rather different estimates of mediation that each suggests, based on their goodness-of-fit? With the simulated data based on the PACE trial sample size, which is quite large by the standards of the field, it was surprising that we could not reject any of the other models based on RMSEA of > 0.05 (Table 1). Therefore, in the real world where we don't know the true model, we may need to consider estimates of mediation from several plausible or near-plausible models to provide a fair description of our level of uncertainty. It is worth mentioning that the simplex models did not have adequate fit by the RMSEA criterion when fitted to the PACE data itself (data not shown); it may be that

the simulated data studied here is less complex than actual data, reducing our ability to discern between models.

But RMSEA is just one criterion. If we use the BIC to judge which was the best fitting of the models, with the best model being the one with the smallest BIC, the order from worst to best fit was simplex with lagged *b* paths, simplex with contemporaneous *b* paths, latent growth model, and modified on latent change model (Table 1). The difference in BIC for each successive model was more than two points, suggesting some scope for discriminating differences in fit between the models (Kass & Raftery, 1995).

Comparing the two simplex models, the model with contemporaneous *b* paths fitted better than the model with lagged *b* paths, with the AIC and BIC more than 5 points lower for the contemporaneous model. This was found with the PACE data as well (Goldsmith et al., 2016), suggesting that we should prefer the results from the model with contemporaneous *b* paths even though this model is less theoretically appealing. We also note that the indirect effects were larger in magnitude in the better fitting contemporaneous *b* path model.

Discussion

This tutorial outlines the fitting of longitudinal mediation SEM to simulated clinical trial data. We have demonstrated four different models, each of which could incorporate further variations. It is generally preferred that a theoretically plausible model is chosen up front and mediation effects evaluated in that single model. However, in practice it is often difficult to get strong guidance on model choice through theory and prior evidence. We suggest that considering the consistency of findings across a range of empirically plausible models provides a robust analysis option. These models should also be chosen as much as possible based on theory. Exploring the findings across models provides a type of sensitivity analysis for the mediation parameters of interest. The guidance in this tutorial should be helpful for either approach. The sensitivity analysis fitting all of these models to the

simulated dataset consistently suggested there was at least partial mediation of the effect of

two of the treatments (R1 and R3) with no mediation of the third treatment effect (R2).

These findings were given credence by their robustness over different model types.

 In terms of the different model types, simplex models have the advantage of being

easily interpretable and of providing coefficient estimates dis-attenuated for some particular

forms of measurement error, but they make strong autoregressive structure assumptions

which may be unrealistic. Latent growth models are perhaps the simplest of the models

studied, but their reduction of the mediator and outcome to single rate-of-change variables

may be overly simplistic and does not as rigorously respect the temporal ordering implied by

a mediation hypothesis as compared to some of the other models. The latent growth models

could be extended to allow different slope/growth latent variables for different time-intervals,

so-called piecewise growth curve models. In fact, the latent change model is a special case of

a piecewise latent growth model. Latent change models may be preferred in situations where

change is not expected to be "uniform" between measurements in a longitudinal design

(Lockhart et al., 2011; MacKinnon, 2008). The PACE trial data followed such a trajectory

(Supplemental Material, Figure S3), which is probably a common pattern for mediator and

outcome measures in clinical trials of psychological and behavioral therapies. There are a

couple of reasons for this. Firstly, eligibility criteria for such trials often include thresholds

based on the outcome variables. Participants therefore tend to be at the more severe end of

the spectrum with greater scope for rapid early change. This may then be followed by a

plateau effect as follow-up continues. For example, in the PACE trial, eligibility was based

on clinically important thresholds on the two primary outcome measures (physical function

and fatigue) (White et al., 2011). The baseline mean values for these measures were

therefore on the more severe end of the spectrum as compared to some other groups of CFS

patients and members of the community (Cella & Chalder, 2010; Chalder, Power, &

Wessely, 1996; Hambrook et al., 2011). In this scenario, which occurs often in clinical trial settings, treatment and regression to the mean can both lead to rapid change. Secondly, there is evidence that some participants experience early gains after commencing psychological therapy, such as in the example of depression (Delgadillo et al., 2014). Latent change models may also be more appropriate when a predictor has a different effect on change at different times (Lockhart et al., 2011; MacKinnon, 2008). In trial data as used here, treatment group will be included in longitudinal mediation models and might be expected to lead to different amounts of change at different times post-randomization, i.e. the *a* paths might differ over time. This was what was seen in the PACE trial data, where there was a shrinking treatment effect on the mediator (*a* path) over time (Goldsmith et al., 2016).

If the reader chooses to apply a specific model rather than take a sensitivity approach, this could be done based on hypotheses about learning in treatments and interventions. For example, CBT teaches people coping skills that they need to remember and apply after treatment ceases. If we believe on average people maintain and continue to apply skills they have learned after treatment finishes leading to continuing improvements in outcomes, a latent growth model might be appropriate as it will model a trajectory that continues in a given direction. Alternatively, we might believe there is a large improvement in the mediator at the beginning of treatment, but some "fallback" after treatment ends when people miss the therapist support or fail to practice some of what they have learned over time. In this case, latent change models might be more appropriate because they allow for variation in an individual's trajectory of change direction during different time periods. If we think there might be a plateau in the treatment effect, we could either try fitting a latent growth model with an appropriate trajectory, for example [0, 1, 1, 1] or piecewise latent growth/latent change model. The additional benefit of fitting a latent change model (or the modified form) is that rather than having to specify a trajectory, these models will more flexibly estimate the

latent change score. This would seem sensible if we have a sufficient number of

measurements and are less certain of our participants' mediator and outcome trajectories.

Recommendations for trial design when longitudinal mediation is of interest

Mediation analysis requires longitudinal data because it studies a causal relationship

with a temporal ordering. Mediation hypotheses, plans to measure the variables of interest,

and an appropriate measurement schedule should all be in place at the design stage of a

clinical trial. Consideration of the appropriate time lags for measurement is an important and

challenging aspect of studying mediation. Wherever possible, studies of mediation in large

randomized trials should be prefaced with smaller pilot studies that clarify the time course of

change of mediator and outcome. These might suggest different measurement schedules than

those generally applied in clinical trials. Whether we have these data or not, it is probably

best to take measures of both mediator and outcome at multiple follow-up time points in a

larger randomized study, so we can gain information on longitudinal mechanisms of action.

For most of the complex models studied in this manuscript, at least three measures are needed

(for example, baseline and two post-randomization measurements), with more flexibility to

explore assumptions afforded by additional repeated measurements.

Model assumptions and limitations

It is important to note the assumptions these models make. Two important ones were

(a) no systematic measurement error over time (which could be allowed for by measurement

error covariances between time points), and (b) that there was measurement invariance over

time in these models (Millsap, 2011), in other words that the instruments are measuring the

same dimension on the same scale at each time point. We have used this tutorial to provide a

starting point for working with this range of models as opposed to a comprehensive

assessment and evaluation of assumptions and psychometric properties of the instruments used to measure mediator and outcome. In addition, we did not have enough measurements to fully evaluate these assumptions. Where possible, readers should assess the plausibility of these assumptions, and we refer them to relevant literature (Millsap, 2011; Newsom, 2015). Evaluation of the plausibility of these and other assumptions is also an interesting area for future work. Taking more measures of mediators and outcomes over time, or taking different measures of these variables at each time point could lead to models that are identified while making fewer assumptions. Using different measures could also facilitate further exploration of sources of measurement error in mediator and outcome variables of interest. In this vein, it is also important to note that these models do not allow for modeling continuous time. While we do not often have measures of the mediator and outcome in continuous time in psychological and program evaluations to date, such data may become more common in future due to advances like wearable technologies. Methods for continuous time mediation modeling are being developed (Deboeck & Preacher, 2016), and would need to be considered in mediation applications where continuous time measurements are available.

Some of the other general assumptions made when fitting mediation models using SEM may have been strong. In particular, in order to simplify the models for the purpose of the tutorial we did not discuss the inclusion of potential baseline confounders and predictors of missing data. However, we would generally recommend these be included in such models in order to adjust for confounding and increase the plausibility of the missing at random assumption (Dunn et al., 2013; MacKinnon & Pirlott, 2015). This being said, the baseline measurements of mediator and outcome included in these longitudinal models may be among the most important confounders (Pickles et al., 2015), suggesting potentially less need for inclusion of additional variables in this longitudinal context.

A note on studying mediation where there is no treatment effect

It has been argued in the past that mediation analysis should not be pursued when the total effect of treatment is not significant. However, it is now generally accepted that if mediation is of interest an analysis should be done regardless of whether there is a significant total effect (Emsley et al., 2010; Goldsmith et al., 2016; MacKinnon, 2008; MacKinnon & Dwyer, 1993; O'Rourke & MacKinnon, 2015; Pek & Hoyle, 2016; Shrout & Bolger, 2002). In this case, rather than focusing on the mediator as a mechanistic variable that explains some part of the total effect of R -> Y (Baron & Kenny, 1986), it is more relevant to focus on the R -> M and M -> Y relationships to see where the theory breaks down (MacKinnon et al., 2002; Pek & Hoyle, 2016). Estimation of the *a* and *b* paths in this situation could clarify why a treatment didn't work, i.e. the treatment may not have been affecting the mediator as hypothesized (the *a* path is not significant), the mediator may not be associated with the outcome (the *b* path is not significant), or the indirect and residual direct effects may have cancelled each other out. Returning to the *a* path as the action theory, and the *b* path as the conceptual theory, the idea of treatment effect mediation is that our treatment has been designed to take action (affect a mediator) that we strongly suspect will have an effect in turn on the outcome, i.e. which has a demonstrated relationship with the outcome in the literature (conceptual theory). A non-significant *a* path could suggest flaws in the action theory, in other words our treatment is not affecting its intermediate targets, whereas a non-significant *b* path could mean that the conceptual theory relationship postulated between mediator and outcome doesn't exist, perhaps at least in the sample being studied. The information we get from a mediation analysis allows us to evaluate these theoretical relationships, with longitudinal models providing more information about these relationships over time. Understanding whether the issue lies in the action theory, the conceptual theory, or both should allow for treatment mechanisms to be revisited, and for treatments to be effectively

refined. For example, if we had expected the R2 treatment (corresponding to the APT treatment in PACE) to affect the given mediator, finding that it has not suggests the treatment may need to be refined. If we could change the treatment such that it has more of an effect on the mediator, which in this case we know to have a significant conceptual theory relationship with the outcome, the treatment might be more effective. We note in the case where there is no effect of the treatment on the outcome, terms like total effect, indirect effect, partial mediation and so on, become rather obsolete. Reverting to this sort of discussion of the *a* path/action theory and *b* path/conceptual theory would seem sensible in this case.

In conclusion

This tutorial discusses the practical application of simplex, latent growth, and latent change models to longitudinal data to study mediation. Modified latent change models are probably the most flexible, and seem promising models to explore when studying mediation in the clinical trials context. In any case, the researcher should decide up front whether to apply a single most theoretically plausible type of model, or to fit a few plausible models, and then state their chosen procedure clearly when publishing results. Investigations further characterizing the effects of the assumptions made in these models, possibly by assessing the tradeoff between plausibility of the assumptions made in more parsimonious models versus gathering more repeated measurements, would allow this tutorial to be usefully updated in future.

References

Akaike, H. (1974). New look at statistical model identification. *Ieee Transactions on Automatic Control, 19*(6), 716-723. doi: 10.1109/Tac.1974.1100705

Bagozzi, R. P., & Heatherton, T. F. (1994). A general approach to representing multifaceted personality constructs: Application to state self-esteem. *Structural Equation Modeling, 1*(1), 35-67. doi: 10.1080/10705519409539961

Baguley, T. (2009). Standardized or simple effect size: What should be reported? *British Journal of Psychology, 100*, 603-617. doi: 10.1348/000712608X377117

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*(6), 1173-1182.

Biesanz, J. C., Deeb-Sossa, N., Papadakis, A. A., Bollen, K. A., & Curran, P. J. (2004). The role of coding time in estimating and interpreting growth curve models. *Psychological Methods, 9*(1), 30-52. doi: 10.1037/1082-989X.9.1.30

Bollen, K. A. (1989). *Structural equations with latent variables*. New York, NY: John Wiley & Sons.

Bollen, K. A., & Noble, M. D. (2011). Structural equation models and the quantification of behavior. *Proceedings of the National Academy of Sciences United States of America, 108*, 15639-15646. doi: 10.1073/pnas.1010661108

Browne, M. W., & Cudeck, R. (1992). Alternative ways of assessing model fit. *Sociological Methods & Research, 21*(2), 230-258. doi: 10.1177/0049124192021002005

Buchwald, D., Pearlman, T., Umali, J., Schmaling, K., & Katon, W. (1996). Functional status in patients with chronic fatigue syndrome, other fatiguing illnesses, and healthy individuals. *American Journal of Medicine, 101*(4), 364-370.

Cella, M., & Chalder, T. (2010). Measuring fatigue in clinical and community settings. *Journal of Psychosomatic Research, 69*(1), 17-22. doi: 10.1016/j.jpsychores.2009.10.007

Chalder, T., Goldsmith, K. A., White, P. D., Sharpe, M., & Pickles, A. (2015). Rehabilitative therapies for chronic fatigue syndrome: a secondary mediation analysis of the PACE trial. *Lancet Psychiatry, 2*(2), 141-152. doi: 10.1016/S2215-0366(14)00069-8

Chalder, T., Power, M. J., & Wessely, S. (1996). Chronic fatigue in the community: 'A question of attribution'. *Psychological Medicine, 26*(4), 791-800.

Chen, H. (1990). Action theory and conceptual theory: summatively diagnosing the intervention program. *Theory-driven evaluations* (pp. 197-204). Newbury Park, CA, USA: Sage Publications, Inc.

Cheong, J., MacKinnon, D. P., & Khoo, S. T. (2003). Investigation of mediational processes using parallel process latent growth curve modeling. *Structural Equation Modeling, 10*(2), 238-262. doi: 10.1207/S15328007sem1002_5

Cheung, M. W. L. (2009). Comparison of methods for constructing confidence intervals of standardized indirect effects. *Behavior Research Methods, 41*(2), 425-438.

Cole, D. A., & Maxwell, S. E. (2003). Testing mediational models with longitudinal data: Questions and tips in the use of structural equation modeling. *Journal of Abnormal Psychology, 112*(4), 558-577. doi: 10.1037/0021-843x.112.4.558

De Stavola, B. L., Daniel, R. M., Ploubidis, G. B., & Micali, N. (2015). Mediation analysis with intermediate confounding: Structural equation modeling viewed through the causal inference lens. *American Journal of Epidemiology, 181*(1), 64-80. doi: 10.1093/aje/kwu239

Deboeck, P. R., & Preacher, K. J. (2016). No need to be discrete: A method for continuous

    time mediation analysis. *Structural Equation Modeling, 23*(1), 61-75. doi:

    10.1080/10705511.2014.973960

Delgadillo, J., McMillan, D., Lucock, M., Leach, C., Ali, S., & Gilbody, S. (2014). Early

    changes, attrition, and dose-response in low intensity psychological interventions.

    British Journal of Clinical Psychology, 53(1), 114-130. doi: 10.1111/Bjc.12031

Dunn, G., Emsley, R., Liu, H. H., & Landau, S. (2013). Integrating biomarker information

    within trials to evaluate treatment mechanisms and efficacy for personalised

    medicine. *Clinical Trials, 10*(5), 709-719. doi: 10.1177/1740774513499651

Dunn, G., Everitt, B., & Pickles, A. (1993). *Modelling covariances and latent variables using*

    EQS. London, UK: Chapman & Hall.

Efron, B., & Tibshirani, R. (1993). *An introduction to the bootstrap*. Boca Raton, FL:

    Chapman and Hall/CRC.

Emsley, R., Dunn, G., & White, I. R. (2010). Mediation and moderation of treatment effects

    in randomised controlled trials of complex interventions. *Statistical Methods in*

    Medical Research, 19(3), 237-270. doi: 10.1177/0962280209105014

Enders, C. K., & Bandalos, D. L. (2001). The relative performance of full information

    maximum likelihood estimation for missing data in structural equation models.

    Structural Equation Modeling, 8(3), 430-457. doi: 10.1207/S15328007sem0803_5

Fritz, M. S., Taylor, A. B., & MacKinnon, D. P. (2012). Explanation of two anomalous

    results in statistical mediation analysis. *Multivariate Behavioral Research, 47*(1), 61-

    87. doi: 10.1080/00273171.2012.640596

Goldsmith, K. A., Chalder, T. C., White, P. D., Sharpe, M., & Pickles, A. (2016).

    Measurement error, time lag, unmeasured confounding: considerations for

    longitudinal estimation of the mediation '*b* path' in randomised clinical trials.

Statistical Methods in Medical Research, Advance online publication. doi: 10.1177/0962280216666111

Gollob, H. F., & Reichardt, C. S. (1987). Taking account of time lags in causal models. *Child Development, 58*(1), 80-92. doi: 10.1111/j.1467-8624.1987.tb03492.x

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.

Hambrook, D., Oldershaw, A., Rimes, K., Schmidt, U., Tchanturia, K., Treasure, J., . . . Chalder, T. (2011). Emotional expression, self-silencing, and distress tolerance in anorexia nervosa and chronic fatigue syndrome. *British Journal of Clinical Psychology, 50*, 310-325. doi: 10.1348/014466510X519215

Hedeker, D., & Gibbons, R. D. (1997). Application of random-effects pattern-mixture models for missing data in longitudinal studies. *Psychological Methods, 2*(1), 64-78. doi: 10.1037/1082-989x.2.1.64

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*(1), 1-55. doi: 10.1080/10705519909540118

Imai, K., Keele, L., & Tingley, D. (2010). A general approach to causal mediation analysis. *Psychological Methods, 15*(4), 309-334. doi: 10.1037/A0020761

Jöreskog, K. G. (1970). Estimation and testing of simplex models. *British Journal of Mathematical & Statistical Psychology, 23*(2), 121-145.

Jöreskog, K. G. (1979). Statistical estimation of structural models in longitudinal-developmental investigations. In J. R. Nesselroade & P. B. Baltes (Eds.), *Longitudinal*

research in the study of behavior and development (pp. 303-351). New York, NY: Academic Press, Inc.

Judd, C. M., & Kenny, D. A. (1981). Process analysis - Estimating mediation in treatment evaluations. *Evaluation Review, 5*(5), 602-619.

Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association, 90*(430), 773-795. doi: 10.1080/01621459.1995.10476572

Kline, R. B. (2011). *Principles and practice of structural equation modelling*. New York, NY: The Guildford Press.

Knudsen, A. K., Henderson, M., Harvey, S. B., & Chalder, T. (2011). Long-term sickness absence among patients with chronic fatigue syndrome. *British Journal of Psychiatry, 199*(5), 430-431. doi: 10.1192/bjp.bp.110.082974

Krull, J. L., Cheong, J., Fritz, M. S., & MacKinnon, D. P. (2016). Moderation and mediation in interindividual longitudinal analysis. In D. Cicchetti (Ed.), *Developmental psychopathology* (pp. 922-985). New York, NY: Wiley.

Lockhart, G., MacKinnon, D. P., & Ohlrich, V. (2011). Mediation analysis in psychosomatic medicine research. *Psychosomatic Medicine, 73*(1), 29-43. doi: 10.1097/Psy.0b013e318200a54b

MacKinnon, D. P. (2008). *Introduction to statistical mediation analysis*. New York, NY: Taylor & Francis Group LLC.

MacKinnon, D. P., & Dwyer, J. H. (1993). Estimating mediated effects in prevention studies. *Evaluation Review, 17*(2), 144-158.

MacKinnon, D. P., Lockwood, C. M., Hoffman, J. M., West, S. G., & Sheets, V. (2002). A comparison of methods to test mediation and other intervening variable effects. *Psychological Methods, 7*(1), 83-104. doi: 10.1037//1082-989X.7.1.83

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(1), 99-128. doi: 10.1207/s15327906mbr3901_4

MacKinnon, D. P., & Pirlott, A. G. (2015). Statistical approaches for enhancing causal

     interpretation of the M to Y relation in mediation analysis. *Personality and Social*

     Psychology Review, 19(1), 30-43. doi: 10.1177/1088868314542878

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(2), 157-183. doi: 10.1111/j.1745-3984.1993.tb01072.x

Marsh, H. W., & Hau, K. T. (1996). Assessing goodness of fit: Is parsimony always

     desirable? *Journal of Experimental Education, 64*(4), 364-390.

Maxwell, S. E., & Cole, D. A. (2007). Bias in cross-sectional analyses of longitudinal

     mediation. *Psychological Methods, 12*(1), 23-44. doi: 10.1037/1082-989X.12.1.23

Maxwell, S. E., Cole, D. A., & Mitchell, M. A. (2011). Bias in cross-sectional analyses of

     longitudinal mediation: Partial and complete mediation under an autoregressive

     model. *Multivariate Behavioral Research, 46*(5), 816-841. doi:

     10.1080/00273171.2011.606716

McArdle, J. J. (2009). Latent variable modeling of differences and changes with longitudinal

     data. *Annual Review of Psychology, 60*, 577-605. doi:

     10.1146/annurev.psych.60.110707.163612

McHorney, C. A., Ware, J. E., & Raczek, A. E. (1993). The Mos 36-item short-form health

     survey (SF-36). 2. Psychometric and clinical tests of validity in measuring physical

     and mental health constructs. *Medical Care, 31*(3), 247-263.

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

     Routledge, Taylor & Francis Group.

Moss-Morris, R., & Chalder, T. (2003, September). *Illness representations: Where to from here?* Paper presented at the European Health Psychology Society, Kos, Greece.

Muthen, 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*(4), 371-402. doi: 10.1037//1082-989x.2.4.371

Neale, M. C., & Cardon, L. R. (1992). *Methodology for genetic studies of twins and families.* Dordrecht, Netherlands: Kluwer Academic Publishers.

Newsom, J. (2015). *Longitudinal structural equation modeling: A comprehensive introduction*. New York, NY: Routledge.

O'Rourke, H. P., & MacKinnon, D. P. (2015). When the test of mediation is more powerful than the test of the total effect. *Behavior Research Methods, 47*(2), 424-442. doi: 10.3758/s13428-014-0481-z

Okleshen Peters, C. L., & Enders, C. K. (2002). A primer for the estimation of structural equation models in the presence of missing data: Maximum likelihood algorithms. *Journal of Targeting, Measurement and Analysis for Marketing, 11*(1), 81-95.

Pek, J., & Hoyle, R. H. (2016). On the (in)validity of tests of simple mediation: Threats and solutions. *Social and Personality Psychology Compass, 10*(3), 150-163. doi: 10.1111/spc3.12237

Pickles, A., Harris, V., Green, J., Aldred, C., MacConachie, H., Slonims, V., . . . Charman, T. (2015). Treatment mechanism in the MRC preschool autism communication trial: implications for study design and parent-focussed therapy for children. *Journal of Child Psychology and Psychiatry, 56*(2), 162-170.

Preacher, K. J., & Kelley, K. (2011). Effect size measures for mediation models: Quantitative strategies for communicating indirect effects. *Psychological Methods, 16*(2), 93-115. doi: 10.1037/a0022658

Raykov, T., Marcoulides, G. A., & Boyd, J. (2003). Using SEM programs to perform matrix manipulations and data simulation. *Structural Equation Modeling, 10*(2), 312-322. doi: 10.1207/S15328007sem1002_9

Resseguier, N., Giorgi, R., & Paoletti, X. (2011). Sensitivity analysis when data are missing not-at-random. *Epidemiology, 22*(2), 282. doi: 10.1097/EDE.0b013e318209dec7

Robins, J. M., & Greenland, S. (1992). Identifiability and exchangeability for direct and indirect effects. *Epidemiology, 3*(2), 143-155.

Schwarz, G. (1978). Estimating dimension of a model. *Annals of Statistics, 6*(2), 461-464. doi: 10.1214/aos/1176344136

Selig, J. P., & Preacher, K. J. (2009). Mediation models for longitudinal data in developmental research. *Research in Human Development, 6*(2-3), 144-164. doi: 10.1080/15427600902911247

Shrout, P. E., & Bolger, N. (2002). Mediation in experimental and nonexperimental studies: New procedures and recommendations. *Psychological Methods, 7*(4), 422-445. doi: 10.1037//1082-989x.7.4.422

Skerrett, T. N., & Moss-Morris, R. (2006). Fatigue and social impairment in multiple sclerosis: the role of patients' cognitive and behavioral responses to their symptoms. *J Psychosom Res, 61*(5), 587-593. doi: 10.1016/j.jpsychores.2006.04.018

Sobel, M. (1986). Some new results on indirect effects and their standard errors in covariance structure analysis. In N. B. Tuma (Ed.), *Sociological methodology 1986* (pp. 159-186). Washington, D.C.: American Sociological Association.

Steiger, J. H. (1990). Structural model evaluation and modification - An interval estimation

   approach. *Multivariate Behavioral Research, 25*(2), 173-180. doi:

   10.1207/s15327906mbr2502_4

Steiger, J. H., & Lind, J. C. (1980, May). *Statistically based tests for the number of common

   factors*. Paper presented at the annual spring meeting of the Psychometric Society,

   Iowa City, IA.

Stephenson, M. T., & Holbert, R. L. (2003). A Monte Carlo simulation of observable versus

   latent variable structural equation modeling techniques. *Communication Research,

   30*(3), 332-354. doi: 10.1177/0093650203252410

Steyer, R., Eid, M., & Schwenkmezger, P. (1997). Modeling true intraindividual change:

   True change as a latent variable. *Methods of Psychological Research - Online, 2*(1),

   21-33.

VanderWeele, T. J. (2009). Marginal structural models for the estimation of direct and

   indirect effects. *Epidemiology, 20*(1), 18-26. doi: 10.1097/EDE.0b013e31818f69ce

VanderWeele, T. J., & Vansteelandt, S. (2009). Conceptual issues concerning mediation,

   interventions and composition. *Statistics and Its Interface, 2*(4), 457-468.

VanderWeele, T. J., Vansteelandt, S., & Robins, J. M. (2014). Effect decomposition in the

   presence of an exposure-induced mediator-outcome confounder. *Epidemiology, 25*(2),

   300-306. doi: 10.1097/Ede.0000000000000034

Wang, L., Zhang, Z., & Estabrook, R. (2009). Longitudinal mediation analysis of training

   intervention effects. In S. Chow, E. Ferrer & F. Hsieh (Eds.), *Statistical methods for

   modeling human dynamics: An interdisciplinary dialogue* (pp. 349-380). New York,

   NY: Taylor & Francis.

White, P. D., Goldsmith, K. A., Johnson, A. L., Potts, L., Walwyn, R., DeCesare, J. C., . . .

   Sharpe, M. (2011). Comparison of adaptive pacing therapy, cognitive behaviour

therapy, graded exercise therapy, and specialist medical care for chronic fatigue syndrome (PACE): a randomised trial. *Lancet, 377*(9768), 823-836. doi: 10.1016/S0140-6736(11)60096-2

White, P. D., Sharpe, M. C., Chalder, T., DeCesare, J. C., & Walwyn, R. (2007). Protocol for the PACE trial: a randomised controlled trial of adaptive pacing, cognitive behaviour therapy, and graded exercise, as supplements to standardised specialist medical care versus standardised specialist medical care alone for patients with the chronic fatigue syndrome/myalgic encephalomyelitis or encephalopathy. *BMC Neurology, 7*, 6. doi: 10.1186/1471-2377-7-6

Willett, J. B., & Sayer, A. G. (1994). Using covariance structure analysis to detect correlates and predictors of individual change over time. *Psychological Bulletin, 116*(2), 363-381. doi: 10.1037/0033-2909.116.2.363

Wright, S. (1920a). Correlation and causation Part 1. Method of path coefficients. *Journal of Agricultural Research, 20*, 0557-0585.

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

Zhang, Z., Wang, L., & Tong, X. (2015). Mediation analysis with missing data through multiple imputation and bootstrap. In L. van der Ark, D. Bolt, W. Wang, J. Douglas & S. Chow (Eds.), *Quantitative psychology research* (pp. 341-355). Switzerland: Springer.

Table 1 *Comparison of fit indices in models fitted to simulated data across longitudinal mediation model types*

Model	χ^2	RMSEA [90% CI]	AIC	BIC	AIC difference	BIC difference
Simplex lagged Figure 1	63.3 df = 28 p < 0.001	0.044 [0.030, 0.059] p = 0.72	13479	13658	---	---
Simplex contemporaneous Figure 2	58.0 df =29 p = 0.001	0.040 [0.024, 0.054] p = 0.87	13472	13646	-7	-12
Latent growth Figure 3	65.0 df =33 p < 0.001	0.039 [0.025, 0.053] p = 0.90	13471	13627	-8	-31
Modified latent change Figure 4	26.2 df =33 p = 0.79	0.000 [0.000, 0.020] p > 0.99	13432	13588	-47	-70

Note. χ^2 = chi-square model fit statistic, RMSEA = root mean square error of approximation, CI = confidence interval, AIC = Akaike's Information Criterion, BIC = Bayesian information criterion, df = degrees of freedom, AIC and BIC differences relative to the simplex lagged model.
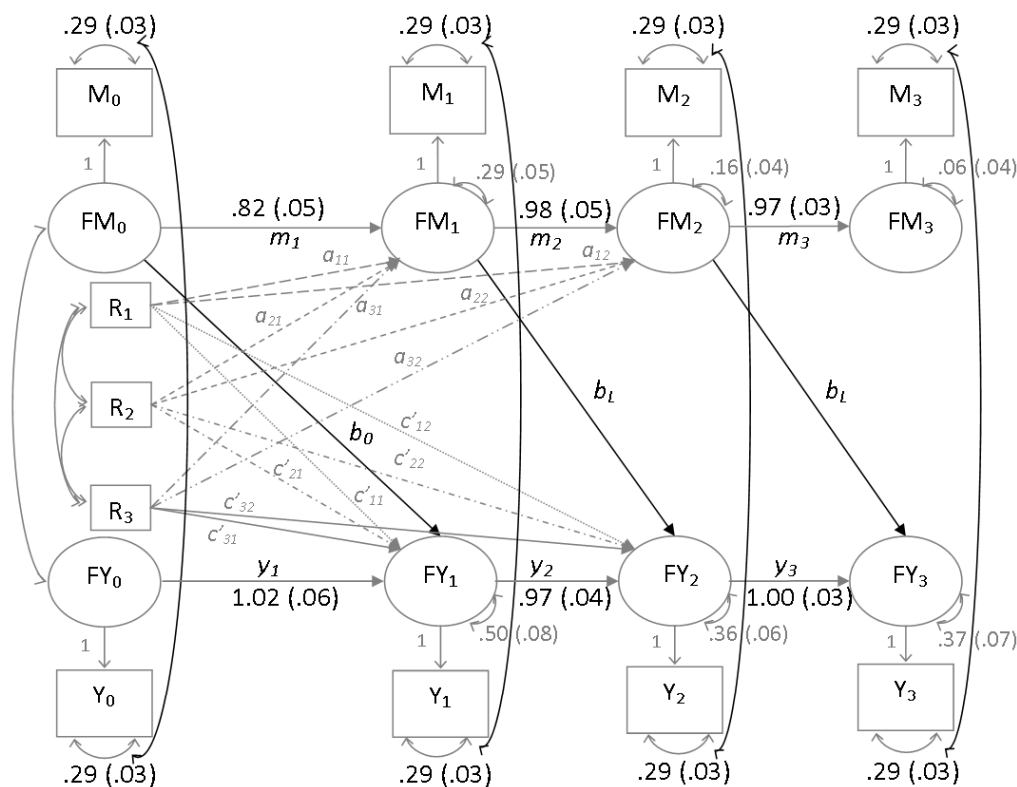
| a_{11}	-.82 [-1.02, -.63]	a_{12}	-.01 [-.20, .18]	c'_{11}	.60 [.40, .81]	c'_{12}	.13 [-.09, .33]
a_{21}	.06 [-.12, .23]	a_{22}	.04 [-.14, .20]	c'_{21}	.02 [-.20, .25]	c'_{22}	.12 [-.10, .32]
a_{31}	-.96 [-1.17, -.79]	a_{32}	.04 [-.15, .25]	c'_{31}	.48 [.26, .71]	c'_{32}	.19 [-.04, .41]
b_0	-.12 [-.25, -.01]	b_L	-.05 [-.10, -.001]				

Paths and effects for third time point	Parameters	R1	R2	R3
$R \to fm_1 \to fy_2 \to fy_3 \to y_3$	$a_{(r\#)1} \times b_L \times y_3 \times 1$	0.04 [0.001, 0.08]	-0.003 [-0.01, 0.01]	0.04 [0.001, 0.09]
$R \to fm_1 \to fm_2 \to fy_3 \to y_3$	$a_{(r\#)1} \times m_2 \times b_L \times 1$	0.04 [0.001, 0.08]	-0.003 [-0.01, 0.01]	0.04 [0.001, 0.09]
$R \to fm_2 \to fy_3 \to y_3$	$a_{(r\#)2} \times b_L \times 1$	0.0005[a] [-0.01, 0.01]	-0.002 [-0.01, 0.01]	-0.002 [-0.01, 0.01]
Overall indirect effect		0.08 [0.001, 0.16]	-0.01 [-0.03, 0.01]	0.09 [0.002, 0.18]
$R \to fy_1 \to fy_2 \to fy_3 \to y_3$	$c'_{(r\#)1} \times y_2 \times y_3 \times 1$	0.58 [0.39, 0.78]	0.02 [-0.18, 0.23]	0.47 [0.25, 0.69]
$R \to fy_2 \to fy_3 \to y_3$	$c'_{(r\#)2} \times y_3 \times 1$	0.13 [-0.09, 0.33]	0.12 [-0.09, 0.32]	0.19 [-0.04, 0.41]
Overall direct effect		0.71 [0.46, 0.97]	0.15 [-0.11, 0.41]	0.66 [0.39, 0.93]
Total effect		0.79 [0.55, 1.04]	0.14 [-0.12, 0.41]	0.74 [0.49, 1.01]

Figure 1. Four group dual process simplex model with lagged *b* paths and contemporaneous residual covariance paths. Numbers in round brackets are standard errors, numbers in square brackets are 95% confidence intervals. The lower table shows indirect and direct effect estimates for the third post-randomization time point. Significant effects shown in bold font, R_1 R_2 and R_3 = dummy variables for randomized treatment group, M_0, M_1, M_2, M_3 =

mediator measurements taken at baseline, 1st follow-up time point, 2nd follow-up time point and 3rd follow-up time point post-randomization, Y_0, Y_1, Y_2, Y_3 = outcome measurements taken at the same time points, FM_0, FM_1, FM_2, FM_3 = true latent mediator scores at the given time points, FY_0, FY_1, FY_2, FY_3 = true latent outcome scores at the given time points, b_0 = "b path" from baseline measure, b_L = lagged b path, m_1, m_2, m_3 = paths between M_0 and M_1, M_1 and M_2, M_2 and M_3, respectively, with y_1, y_2, y_3 the same for the outcome variable, *(r#)* in the table indicates that the number of the treatment group of interest (R1, R2 or R3) should be substituted.

[a]In the Mplus output, this value shows up as 0.000 due to the number of decimal points displayed. It was obtained by creating another parameter multiplied by 100, which is not in the Mplus code included with the tutorial.
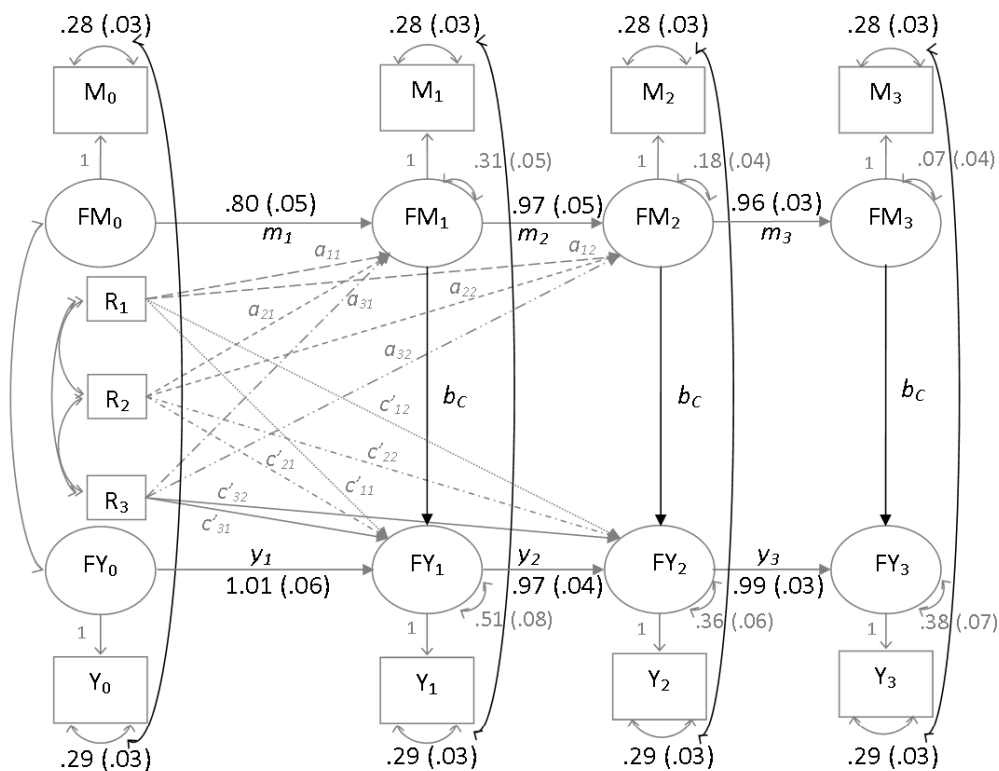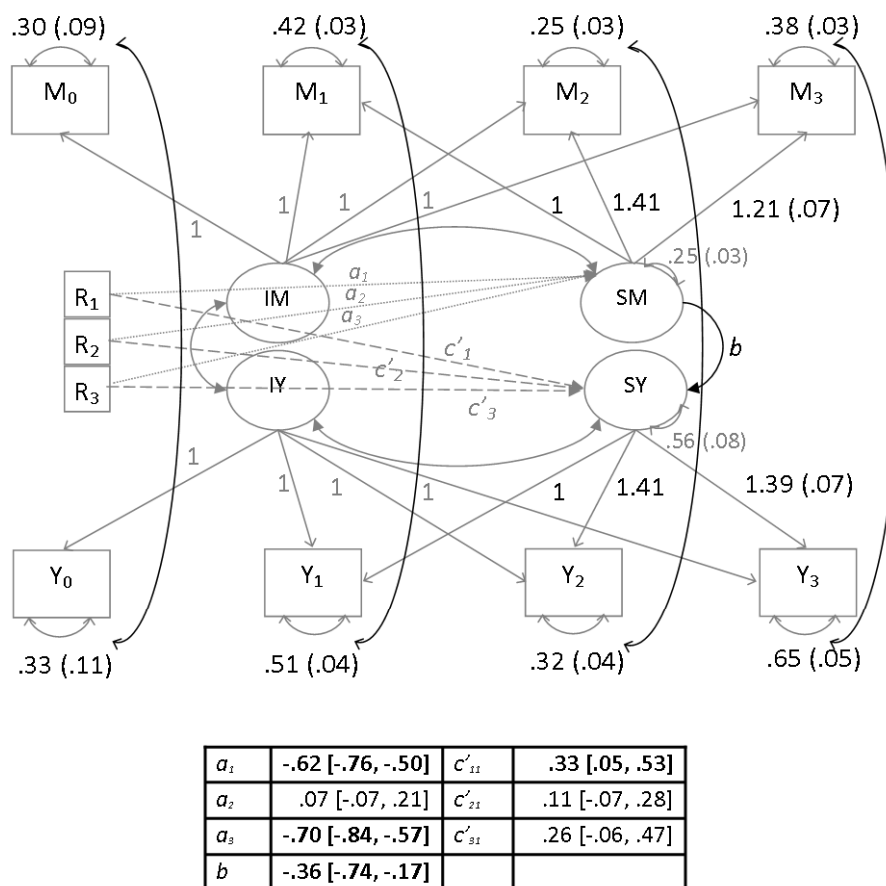
a_{11}	-.82 [-1.01, -.63]	a_{12}	-.02 [-.21, .18]	c'_{11}	.54 [.34, .73]	c'_{12}	.10 [-.11, .29]
a_{21}	.06 [-.12, .23]	a_{22}	.04 [-.14, .21]	c'_{21}	.02 [-.19, .24]	c'_{22}	.13 [-.09, .32]
a_{31}	-.96 [-1.17, -.79]	a_{32}	.03 [-.16, .24]	c'_{31}	.41 [.19, .63]	c'_{32}	.16 [-.05, .38]
b_C	-.08 [-.12, -.04]						

Paths and effects for third time point	Parameters	R1 Estimate	R2 Estimate	R3 Estimate
R->fm_1-> fy_1->fy_2->fy_3 -> y_3	$a_{(\#)1} \times b_C \times y_2 \times y_3 \times 1$	**0.06 [0.03, 0.09]**	-0.005 [-0.02, 0.01]	**0.07 [0.04, 0.11]**
R->fm_1->fm_2->fy_2->fy_3 -> y_3	$a_{(\#)1} \times m_2 \times b_C \times y_3 \times 1$	**0.06 [0.03, 0.10]**	-0.005 [-0.02, 0.01]	**0.07 [0.04, 0.11]**
R->fm_1->fm_2->fm_3->fy_3 -> y_3	$a_{(\#)1} \times m_2 \times m_3 \times b_C \times 1$	**0.06 [0.03, 0.09]**	-0.005 [-0.02, 0.01]	**0.07 [0.04, 0.11]**
R->fm_2->fy_2->fy_3 -> y_3	$a_{(\#)2} \times b_C \times y_3 \times 1$	0.001 [-0.01, 0.02]	-0.003 [-0.02, 0.01]	-0.002 [-0.02, 0.01]
R->fm_2->fm_3->fy_3 -> y_3	$a_{(\#)2} \times m_3 \times b_C \times 1$	0.001 [-0.01, 0.02]	-0.003 [-0.02, 0.01]	-0.002 [-0.02, 0.01]
Overall indirect effect		**0.19 [0.09, 0.28]**	-0.02 [-0.06, 0.02]	**0.21 [0.11, 0.32]**
R -> fy_1 -> fy_2 -> fy_3 -> y_3	$c'_{(\#)1} \times y_2 \times y_3 \times 1$	**0.52 [0.32, 0.71]**	0.02 [-0.18, 0.23]	**0.39 [0.18, 0.61]**
R -> fy_2 -> fy_3 -> y_3	$c'_{(\#)2} \times y_3 \times 1$	0.10 [-0.11, 0.29]	0.13 [-0.08, 0.33]	0.16 [-0.06, 0.37]
Overall direct effect		**0.62 [0.37, 0.87]**	0.15 [-0.10, 0.41]	**0.55 [0.28, 0.81]**
Total effect		**0.80 [0.56, 1.06]**	0.13 [-0.12, 0.41]	**0.76 [0.51, 1.03]**

Figure 2. Four group dual process simplex model with contemporaneous *b* paths and contemporaneous residual covariance paths. Numbers in round brackets are standard errors, numbers in square brackets are 95% confidence intervals. The lower table shows indirect and direct effect estimates for the third post-randomization time point. Significant effects shown in bold font, R_1 R_2 and R_3 = dummy variables for randomized treatment group, M_0, M_1, M_2,

M_3 = mediator measurements taken at baseline, 1st follow-up time point, 2nd follow-up time point and 3rd follow-up time point post-randomization, Y_0, Y_1, Y_2, Y_3 = outcome measurements taken at the same time points, FM_0, FM_1, FM_2, FM_3 = true latent mediator scores at the given time points, FY_0, FY_1, FY_2, FY_3 = true latent outcome scores at the given time points, b_0 = "b path" from baseline measure, b_C = contemporaneous b path, m_1, m_2, m_3 = paths between M_0 and M_1, M_1 and M_2, M_2 and M_3, respectively, with y_1, y_2, y_3 the same for the outcome variable, *(r#)* in the table indicates that the number of the treatment group of interest (R1, R2 or R3) should be substituted.

		c'_{11}	.33 [.05, .53]
a_1	-.62 [-.76, -.50]	c'_{11}	.33 [.05, .53]
a_2	.07 [-.07, .21]	c'_{21}	.11 [-.07, .28]
a_3	-.70 [-.84, -.57]	c'_{31}	.26 [-.06, .47]
b	-.36 [-.74, -.17]		

Paths and effects for third time point	Parameters	R1 Estimate	R2 Estimate	R3 Estimate
Overall indirect effect				
R -> SM -> SY -> y_3	$a_{(r\#)}$ x b x 1.39	**0.31 [0.14, 0.63]**	-0.04 [-0.13, 0.04]	**0.35 [0.16, 0.73]**
Overall direct effect				
R -> SY -> y_3	$c'_{(r\#)}$ x 1.39	**0.46 [0.07, 0.75]**	0.15 [-0.09, 0.41]	0.35 [-0.08, 0.67]
Total effect		**0.77 [0.54, 1.03]**	0.12 [-0.13, 0.38]	**0.71 [0.46, 0.97]**

Figure 3. Four group dual process latent growth model, square root of time point slope loadings, final loading estimated, with contemporaneous residual covariance paths. Numbers in round brackets are standard errors, numbers in square brackets are 95% confidence intervals. The lower table shows indirect and direct effect estimates for the third post-randomization time point. Significant effects shown in bold font, R_1 R_2 and R_3 = dummy variables for randomized treatment group, M_0, M_1, M_2, M_3 = mediator measurements taken at baseline, 1st follow-up time point, 2nd follow-up time point and 3rd follow-up time point post-randomization, Y_0, Y_1, Y_2, Y_3 = outcome measurements taken at the same time points, IM =

intercept for the mediator, SM = slope for the mediator, IY = intercept for the outcome, SY =

slope for the outcome, covariances are allowed between IM and SY and IY and SM in the

model but are not shown in the figure, *(r#)* in the table indicates that the number of the

treatment group of interest (R1, R2 or R3) should be substituted.
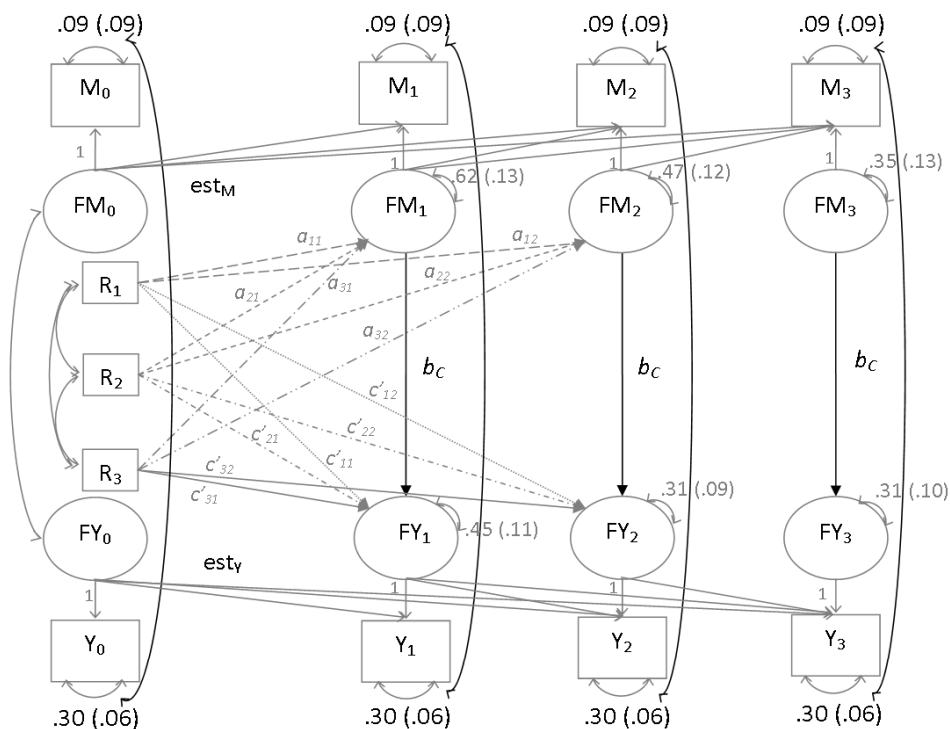
a_{11}	-.83 [-1.01, -.63]	a_{12}	-.31 [-.49, -.10]	c'_{11}	.34 [.07, .57]	c'_{12}	.05 [-.14, .25]
a_{21}	.06 [-.12, .23]	a_{22}	.07 [-.09, .25]	c'_{21}	.04 [-.18, .26]	c'_{22}	.14 [-.07, .34]
a_{31}	-.95 [-1.16, -.78]	a_{32}	-.36 [-.57, -.11]	c'_{31}	.18 [-.10, .43]	c'_{32}	.12 [-.10, .35]
b_C	-.32 [-.53, -.18]						
est_m	.63 [.51, .76]	est_y	1.02 [.83, 1.18]				

Paths and effects for third time point	Parameters	R1 Estimate	R2 Estimate	R3 Estimate
R -> fm$_1$ -> fy$_1$ -> y$_3$	$a_{(r\#)1}$ x b_C x 1.02	**0.27 [0.15, 0.45]**	-0.02 [-0.08, 0.04]	**0.31 [0.17, 0.51]**
R -> fm$_2$ -> fy$_2$ -> y$_3$	$a_{(r\#)2}$ x b_C x 1.02	**0.10 [0.04, 0.17]**	-0.02 [-0.09, 0.03]	**0.12 [0.04, 0.18]**
Overall indirect effect		**0.37 [0.21, 0.57]**	-0.04 [-0.13, 0.04]	**0.42 [0.26, 0.64]**
R -> fy$_1$ -> y$_3$	$c'_{(r\#)1}$ x 1.02	**0.35 [0.07, 0.60]**	0.04 [-0.18, 0.25]	0.18 [-0.10, 0.46]
R -> fy$_2$ -> y$_3$	$c'_{(r\#)2}$ x 1.02	0.05 [-0.14, 0.25]	0.15 [-0.07, 0.36]	0.12 [-0.10, 0.34]
Overall direct effect		**0.40 [0.09, 0.68]**	0.19 [-0.07, 0.44]	0.31 [-0.01, 0.59]
Total effect		**0.77 [0.53, 1.02]**	0.14 [-0.12, 0.41]	**0.73 [0.47, 0.99]**

Figure 4. Four group dual process modified latent change score model with contemporaneous mediation and residual covariance paths. Numbers in round brackets are standard errors, numbers in square brackets are 95% confidence intervals. The lower table shows indirect and direct effect estimates for the third post-randomization time point. Significant effects shown in bold font, est$_m$ = estimates for all mediator measure factor loadings except at the same time point, which is set = 1 to provide the latent variable scale, est$_y$ = estimates for outcome

measure as described for the mediator, R_1 R_2 and R_3 = dummy variables for randomized

treatment group, M_0, M_1, M_2, M_3 = mediator measurements taken at baseline, 1st follow-up

time point, 2nd follow-up time point and 3rd follow-up time point post-randomization, Y_0, Y_1,

Y_2, Y_3 = outcome measurements taken at the same time points, FM_0 = true latent mediator

score at baseline, FM_1, FM_2, FM_3 = modified true latent mediator change between each time

point and the previous time point, FY_0 = true latent outcome score at baseline FY_1, FY_2, FY_3

= modified true latent outcome change between each time point and the previous time point,

(r#) in the table indicates that the number of the treatment group of interest (R1, R2 or R3)

should be substituted.