16

RETs with Follow-Ups: Longitudinal Modeling

Causality = Correlation + Assumptions

- J. BERKEMIJER

Navigation Tips

On the *Programs* tab of my website, programs are provided to conduct non-Mplus analyses in this chapter. Most programs have video links that briefly explain the method, show how to use the program, and review output. I provide links to the videos throughout this chapter. On the *Resources* tab of my webpage, scroll to Chapter 16 for links to relevant online tutorials. To import data from other stat packages into R format for my programs, use the *Import/Syntax* tab on my website. To learn the basics of Mplus syntax, see *Chapter 11* and the *Syntax* tab on my website. For general Mplus output structure, see the *Output* tab on my website (hover the cursor over output text for tooltips).

*** To open a hyperlink in the same window, click on it. To open it in a new background tab, press Ctrl+click (for Macs, Cmd+click). To bring it to immediate focus in a new tab when you activate it, press Ctrl+shift+click (or Cmd+Shift+click for Macs) on the link. To search for a term within an open pdf or on a web page, use Ctrl+f (for Macs, Cmd+f). ***

INTRODUCTION

This chapter focuses on the analysis of RETs that have a longitudinal focus by incorporating three or more repeated assessments of the same construct(s), such as an outcome and/or a mediator. Usually one of these assessments is a baseline measure and another is an "immediate" posttest measure. In studies that use multi-session interventions spread out across time, researchers sometimes obtain additional assessments during the treatment, say at mid-treatment, often with the idea of identifying early responders to the treatment or to identify people who have little hope of responding to treatment over the long run. For example, if for a 12 session treatment to reduce depression it is evident by mid-treatment (session 6) that a person is not responding to the intervention, the decision might be made to terminate the treatment and shift the patient to a new one. Or, if certain people are early responders, they may not need to complete the full treatment protocol and can terminate participation early. In other studies, additional assessments are made after the immediate posttest to document the persistence or decay of intervention effects over time. There are many variants of longitudinal RETs and I cannot consider all of them here. In addition to trials with follow-ups and mid-treatment assessments, I address the analysis of interventions designed to alter trajectories

or growth curves, interventions to affect time-to-event phenomena (survival analysis), and RETS with intensive longitudinal structures, such as daily diaries, to name a few.

In the current chapter, I begin by reviewing a wide range of concepts that are central to longitudinal RETs. The presentation is eclectic because my focus is on explicating core concepts that lay a foundation for longitudinal modeling. Issues include specifying time lags to use in RET designs; facet analyses of causal inference in longitudinal research; the assumptions of stability, stationarity, and equilibrium; contemporaneous and longitudinal assessments of reciprocal causality; addressing confounds; autocorrelation and autoregression; measurement invariance across time; between-person and within-person analytic frameworks; the traditional cross-lagged panel model; and full versus limited information estimation in longitudinal RETs with many follow-ups.

After consideration of these building blocks, I describe different forms of statistical modeling that have been or can be used for the analysis of longitudinal RET designs. The breath of coverage is considerable and includes SEM-based fixed effects and random effects frameworks, SEM-based latent growth modeling, SEM-based survival analysis, dynamic structural equation models for intensive longitudinal data, SEM-based latent change models, SEM-based mixed-models and SEM based midtreatment and adaptive designs. For each framework, I consider the three fundamental RET questions for mediation analysis, namely (1) does the intervention have a meaningful effect on the outcome, (2) does the intervention meaningfully affect the presumed mediators, and (3) do the presumed mediators meaningfully impact the target outcomes.

Like Chapter 15, the current chapter is long and not amenable to processing in a single sitting. It is almost book length in nature. After reading the initial foundation sections, you can read the remaining sections independently depending on your needs and interests. As I suggested in Chapter 15, a useful mindset might be to treat the chapter more like an encyclopedia where you have specific topics that look intriguing based on a quick scan of the chapter outline; then jump directly to that topic and read the 15 or so pages dedicated to it. The common thread through all topics is their focus on RETs and their use of SEM (either LISEM or FISEM). Having said that, I recommend you make sure to read the introductory sections that lay key foundations.

CHOOSING TIME LAGS AND THE NUMBER OF FOLLOW-UPS

When testing causal models, one must make methodological decisions about the time intervals to use between the assessment of causes and effects. When mapping the effect of the treatment on an outcome or mediator, a key question is how long it takes for treatment exposure to produce meaningful change in the outcome/mediator. If you misspecify this interval then, according to many methodologists, you run the risk of

making faulty causal inferences. For example, if you design your study so that your time interval is too short, then it could appear that your intervention is ineffective because you have not given it the time it needs to work. If, on the other hand, you design your study so that your time interval is too long, then the treatment may have worked but then its effects may have decayed so that by the time you assess the outcome, the treatment appears to be weak and/or ineffective. Ultimately, one must make educated guesses about how long it takes for changes in your presumed causes to produce meaningful change in their presumed effects. Unfortunately, researchers often receive little guidance on how to make time lag choices, relying instead on tradition, intuition, chance, or convenience.

Doorman and Griffen (2015) define an optimal time lag for causal analysis as the time lag that produces the maximum effect of the cause on the effect. This conceptualization places priority on learning about the maximal causal effect between two variables. However, this often is not the goal of program evaluation. Rather, we might focus on time lags that are of interest in their own right. If individuals complete a therapy to reduce clinical depression, I might want to know how effective the program is just after the intervention finishes, three months after the intervention finishes, and then again at 6 months and 12 months after the intervention. I basically set time lags in my study to correspond to what I consider to be meaningful short term follow-ups, moderate term follow-ups, and long term follow-ups. The criteria that define "short," "moderate," and "long" term are context specific and depend on both substantive and practical considerations. For example, I might think about how long it takes for the negative consequences of intervention decay to activate as a result of decay and then use this information as a basis for defining follow-up periods. The choice of time periods based on such criteria might vary depending on the vulnerability (capacity to cope) of the target population. From yet another perspective, many program designers simply want to know how long meaningful effects of the intervention last, with time lags chosen to answer this question. Note that such questions are quite different then trying to determine at what point the effect of a change in X on Y is at its maximum.

A general working assumption of many researchers is that causal effects weaken as the time interval between the cause and measured effect increases. There is literature to support this proposition in some content domains but not others. For example, if you measure people's attitudes towards voting for a certain candidate one day prior to election day, there usually will be a stronger association between the attitude and voting behavior with respect to that candidate than if the attitude is measured, say, six months before election day. An example where the principle does not apply would be an intervention that teaches middle school youth techniques for studying for tests to increase their GPA. It often takes considerable time for students to learn and apply new study

habits that eventually impact their overall GPA. In this case, a more appropriate time lag between the intervention and the outcome might be one or two semesters.

Methodologists have sought to elucidate general principles for selecting time intervals for purposes of causal analysis but I often find the principles do not apply to the areas in which I work, i.e., they are not universal. Timmons and Preacher (2015) offer guidelines for choosing time lags based on a series of simulation studies they conducted. They found that (1) adding measurement occasions often increases the accuracy of characterizing nonlinear change across time, (2) when seeking to find the slope of a linear decay function, extreme spacing is ideal, and (3) for nonlinear decay functions, it is beneficial statistically to concentrate measurement occasions where curvature is thought to be the greatest (see also Zhao et al., 2021). McNeish et al. (2023) offer advice for modeling individual differences in the timing of change onset and offset and offer perspectives on such change that might be useful for some RET designs.

In addition to timing, decisions need to be made about the number of follow-ups to conduct in an RET. This often is constrained by project resources. Too many repeated assessments can cause testing and/or practice effects, it can contribute to attrition either positively or negatively and, in some cases, can make recruitment more difficult. Increasing the number of follow-up assessments often positively affects power and precision for purposes of curve analyses but there usually are diminishing returns for doing so (Timmons & Preacher, 2015). When mapping decay curves researchers often specify over what time interval they want to map the curve (e.g., a year) and then try to obtain assessments during that time interval spaced strategically over defined time points. The use of frequent assessments between two theoretically *a priori* specified time points is sometimes referred to as **shortintudinal** research (Doorman & Griffen, 2015).

In the final analysis, I find it difficult to suggest general principles for time lag selection because criteria vary so much as a function of the content area and the different facets of causal effects. In RETs for program evaluation, the most common focus seems to be on evaluating decay or persistence of effects over short-term follow-ups, moderate-term follow-ups, long-term follow ups, and/or, perhaps, extended follow-ups, with each of these time periods being of interest in their own right. However, how one defines them can vary considerably depending on the substantive area of interest.

THE FACETS OF CAUSAL INFERENCE

Causal effects do not occur in isolation. They occur in certain contexts/settings, at certain points in time and for certain types of people. When characterizing a causal link, I think it best to always keep in mind each of these facets and the degree to which we can generalize across them. Also important is characterizing the strength of the causal effect,

the nature of the causal effect, and the expected duration (or decay) of the effect in light of these facets. For example, some theorists argue that anxiety causes depression. This statement is an oversimplification and ignores how long it takes for anxiety to cause depression, for what populations the effect occurs, in what contexts it occurs, what the nature and strength of the relationship is, and how long the effect lasts.

This point is fundamental. The different facets qualify every causal link in a causal model and questions the notion that there exists a single, abstract causal coefficient between variables that scientists should seek to discover. Gollob and Reichardt (1991) emphasize this point when thinking about time lags for mediation analyses. Consider the mediational model $T_{TIME\ 1} \rightarrow M_{TIME\ 2} \rightarrow Y_{TIME\ 3}$, where T is the treatment condition that is introduced at time 1, M is the presumed mediator that is assessed at time 2, and Y is the presumed outcome that is assessed at time 3, i.e., the three elements of the mediational chain are time sequenced. Gollob and Reichardt describe the above chain as a **time-specific indirect effect** that reflects the degree to which M at exactly Time 2 mediates the effect of T at exactly Time 1 on Y at exactly Time 3. Another term they use to characterize the effect is that of a **time-specific mediation effect**. Suppose an RET mediation model contains a mediational chain with the following links: $T_{TIME\ 1} \rightarrow M_{TIME\ 2} \rightarrow M_{TIME\ 3} \rightarrow Y_{TIME\ 4}$. The time specific indirect effect in this case would be framed as the degree to which M at exactly Time 2 mediates the effect of T at exactly Time 1 on M at exactly Time 3 which, in turn, influences Y at exactly Time 4. To be more precise, it refers to the degree to which M at Time 2 mediates the effect of T at exactly Time 1 on M at exactly Time 3 which, in turn, influences Y at exactly Time 4. All of this is further qualified by the particular population of individuals studied and the context/setting in which T, M and Y take place.

Some researchers argue that the choice of time frames in longitudinal designs for purposes of mediation analysis can bias estimates of mediation effect size if the "wrong" time points are selected. Implicit in such arguments is the idea that there is a "correct" time lag, a position that ignores the multi-faceted character of mediation effects. It is analogous to referencing the reliability of a scale in psychometrics while failing to appreciate that there usually is no such thing as a single reliability coefficient that adequately characterizes scale reliability. Reliability of a scale instead is facet dependent. The reliability of the CES-D depression scale, for example, varies depending on who the scale is applied to (e.g., adolescents versus young adults versus seniors), the context in which it is applied (e.g., whether it is administered face-to-face or is self-administered), the time that it is administered, and so on. Such also is true of mediation analyses per the concept of time specific mediation effects described above in conjunction with the other facets of causal links.

Once you select one or more time points for mediation assessment in your study, then your conclusions about mediation are bound by those selected time points. Suppose I implement an intervention to increase motivation to apply oneself in school and its effect on student GPA. If I measure these variables contemporaneously just after the intervention I might find significant effects of the intervention on motivation but that these effects do not translate into intervention effects on current GPA. A conclusion that changes in motivation are not associated with changes in concurrent GPA correctly characterizes the causal coefficient in this case *given the operative time intervals*. It is not that the chosen time interval is "wrong" for estimating the causal link between changes in motivation and changes in GPA; the causal coefficient has been accurately captured. However, if I introduce, say, a six month time lag between the measurement of motivation on the one hand and student GPA on the other hand, then for this particular time lag, the causal coefficient likely will be non-zero and meaningful in its magnitude. The causal coefficient is tied to the time lag facet of the causal link.

The bottom line is that if you are interested in generalizing beyond the chosen time intervals in your program evaluation, then you need to formulate generalizability research designs that evaluate the consistency of mediation dynamics across a variety of target time points. Preacher (2015) has expressed the idea nicely: "*Typically, there is no one 'correct' lag at which to assess an effect; rather, effects tend to vary as a function of lag, and a more complete understanding of relationships among variables can result from assessing them at multiple lags*." Taking an aspirin for a headache is unlikely to have an effect within two minutes, is likely to have an effect after an hour or so, and is unlikely to have much effect after 24 hours. Each of these statements is a valid characterization of the causal effects of aspirin on headaches. It is not the case that one of them is "correct" and the others are "incorrect." Characterizations of causal links are facet dependent.

STABILITY, STATIONARITY, AND EQUILIBRIUM

Three concepts that I make use of in this chapter when discussing models of longitudinal dynamics are those of stability, stationarity, and equilibrium. **Stability** refers to the process of unchanging levels of a variable over time (Kenny, 1979). For example, physical height typically asymptotes around the age of 20, after which it exhibits stability. The concept of stability also is sometimes used to refer to cases where the ordering of individuals on a variable at one point in time has the same ordering at a second point in time (Preacher, 2015). **Stationarity** refers to an unchanging causal structure over time, i.e. it implies that the strength of a causal effect of X on Y remains the same at different time points. **Equilibrium** refers to a causal system that shows

stability in the patterns of covariances and variances over time. In practice, these terms are used somewhat inconsistently which ultimately can be confusing. Sometimes the properties are important for valid model tests and other times not. I elaborate more on these properties throughout this chapter.

CONTEMPORANEOUS AND LONGITUDINAL CAUSAL INFERENCE

It is not uncommon for methodologists to make statements about the inappropriateness of cross-sectional research designs for testing mediational models. The implication is that only longitudinal designs can be used to meaningfully evaluate mediation. The same is often said about documenting causal effects more generally. Such assertions are untrue or, at the very least, should be stated in more nuanced terms.

The most common objections to studying cause-effect relationships using cross-sectional designs is that one cannot easily rule out reciprocal causality nor spuriousness. Although this is technically true, longitudinal designs suffer the same limitations, i.e., many forms of longitudinal designs cannot easily rule out reciprocal causality nor can they rule out spuriousness. Specialized methodologies and analytic strategies are needed to address such matters in cross-sectional designs but this also is true of longitudinal designs. I elaborate these arguments in more depth throughout this chapter.

A common mistake when characterizing cross-sectional research is the assertion that all the variables on which data are collected reference the same time point. In many cross-sectional studies, researchers include time frames into their measures such that there are clearly demarcated lags between presumed causes and effects. For example, I might be interested in the effect of childhood trauma on current mental health, in which case I would focus my childhood trauma questions on retrospective accounts of such trauma and my mental health questions on current accounts of mental health. In this case, there is indeed a time lag between the variables of interest; it is just that the assessments of them have been made concurrently. Or, if I am studying the possible effect of divorce on current child mental health, I might obtain in the same survey a retrospective report from a parent of when the divorce occurred and measures of current child mental health as reported by parents. A time lag between the variables is explicit.

When this strategy is used, it is important that one use valid and reliable indicators of the constructs of interest and that one can make a good case for the psychometric viability of the measures. There are large literatures on how to construct valid retrospective accounts and creative strategies for validating them (for examples, see Jaccard & Jacoby, 2020; Jaccard, McDonald, Wan, Dittus. & Quinlan, 2002, 2004 ; Jaccard & Wan, 1995). The fact that one can incorporate time intervals into the questions that are asked cross-sectionally makes mediation analysis potentially viable in cross-

sectional designs. If you use multiple indicators of retrospective accounts, this is all the more advantageous because you can then take into account forms of systematic and random measurement error in your modeling.

A scenario where cross-sectional research may be more appropriate than longitudinal research for mediation analysis is when there are very short time lags for how long it takes the cause to translate into the effect. For example, the time interval might be a matter of seconds or minutes. In such cases, it makes little sense to force a questionable time lag onto the research design just to make it longitudinal. If the causal dynamics play themselves out in the short run, a cross-sectional study may be better suited to capturing what has transpired.

In psychology, short time intervals between causes and effects are common, especially for processes of learning, perception, and cognition. The priming of beliefs and attitudes are often central to split-second or last-minute decisions as is the near instantaneous accessing of mental scripts from long term memory which then impact behavior and/or the judgments that people make. Skinner's concept of immediate reinforcement is central to shaping behavior as individuals learn to associate behavioral performance with positive or negative consequences. In the health domain, exposures to an allergen (such as peanuts, bee stings, or latex) can trigger immediate adverse responses on the part of one's immune system as it releases chemicals in reaction to the presence of the allergen. Engaging in strenuous exercise can trigger abrupt malfunction in the heart causing it to stop beating. Events that deprive the brain of oxygen can cause immediate strokes that can have devastating consequences for individuals. All of these constitute very short time lags between cause and effect. In sociology, violating a social norm can lead to immediate negative reactions from people who observe the violations. A single action in a crowd (e.g., shouting "fire") can create immediate non-trivial collective reactions on the part of entire crowds of people. The sociology of finance examines how social phenomena and social media, not just economic fundamentals, drive stock market fluctuations, sometimes virally in a matter of minutes. In economics, the release of economic data by government agencies can cause traders to react within minutes, leading to increased trading volume in stocks and bonds.

Researchers often are put in situations where they simply cannot obtain separate measures of the cause and effect at different time points either because the relevant causal lags are too short, because it is not practical to do so, or because of setting sensitivity. Life satisfaction impacts job satisfaction and vice versa but we usually can't harness the complex micro-level causal interplay as the reciprocal dynamics between these variables unfolds. These types of difficulties do not mean that we can't address issues of reciprocal causation between these variables. To be sure, it can be challenging, but as noted, such

also is true of longitudinal designs when choosing appropriate time intervals at which to obtain one's measures. In the final analysis, sometimes cross-sectional/contemporaneous causal modeling will lend itself best to causal inference, sometimes longitudinal data collection will lend itself best to causal inference, and sometimes a mixture of the two will be best for causal inference.

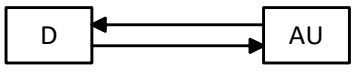
When causal lags are short and if you believe reciprocal causal dynamics are at play, then we need a method for gaining perspectives on the causal dynamics for the case where the measures are obtained cross-sectionally. It turns out, such methods exist and I describe these in the next section. This topic also is important for longitudinal research because, as I show later, longitudinal models sometimes include reciprocal causality between variables not only sequentially across multiple waves but also contemporaneously between variables measured at the same wave.

Addressing "Contemporaneous" Reciprocal Causality

Sometimes we measure variables at a single time point but the current association between them is the result of prior reciprocal causal processes that have already occurred that we are unable to tap into. Perhaps the reciprocal causal dynamics took place seconds, minutes, days, weeks, or even months ago with the net result of these dynamics now reflected in the current correlation between the target variables. Depression often is said to cause heavy alcohol use because drinking alcohol can be a coping strategy for being depressed. By the same token, heavy alcohol use can lead to depression because such use can negatively impact job performance and/or family/peer relations whose deterioration then induces depression. These reciprocal causal dynamics might play themselves out over the course of weeks or a few months in a sequence, like this

$$D_{t1} \rightarrow AU_{t2} \rightarrow D_{t3} \rightarrow AU_{t4}$$

where D represents depression at time *t*, AU represents alcohol use at time *t*, and the numerical subscript attached to *t* represents sequentially later time points as the numbers increase. In a cross-sectional or even in a longitudinal design, we may be unable to assess such processes at this fine-grained level. Our contemporaneous assessments of D and AU, however, may reflect the case where these processes have occurred in the recent past at a more micro-level. The following causal representation captures this possibility:



When we correlate D and AU to explore the possible causal impact of depression on

alcohol use contemporaneously, the correlation can overestimate the causal impact of D on AU because it not only includes the effect of D on AU but it also includes the effect of AU on D. If we want to gain perspectives on the effects of D on AU in light of the reciprocal causal dynamics, we need to adopt specialized analytic strategies.

Suppose I insert the two reciprocal causal paths into a model to apply SEM for purposes of data analysis. In this case, the model will be under-identified and inestimable; there are two unknowns, namely the two path coefficients, but only one known, the single covariance between the variables. However, it turns out that coupled with certain assumptions I might still be able to gain perspectives on the reciprocal causal dynamics in cross-sectional data using an analytic method known as **instrumental variable analysis**, a method I briefly introduced in Chapter 6. Indeed, the "contemporaneous" coefficient estimates might even be better than those obtained by an ill-designed longitudinal study of the same variables if the causal time lag between the variables is short.

To keep my discussion of contemporaneous instrumental variable analysis manageable, the example I will use is a simple, hypothetical one-mediator RET. Suppose a researcher wants to test the effectiveness of a political message to make more positive peoples' attitudes towards supporting a policy more positive by increasing the strength of a specific positive belief about the policy. The belief is conceptualized as the mediator, M, that is thought to influence the attitude, Y. People in the intervention condition are exposed to the message while those in the control group are exposed to an irrelevant message. Just after message exposure, endorsement of the positive belief about the policy is measured for all individuals as is the global attitude toward the policy. The two variables are measured post-exposure on the same questionnaire (i.e., cross-sectionally) based on the assumption that the newly induced variations in the belief translate almost instantly into newly induced variations in the policy attitude. However, suppose the researcher recognizes the possibility of a reciprocal causal dynamic, namely, that one's revised global attitude toward supporting the policy might impact endorsement of the positive belief via rationalization-like cognitive processes; that is, people may be more likely to endorse the positive belief because they have a positive attitude towards the policy and want to justify that attitude accordingly. Thus, the belief may impact the attitude but the attitude also "feedbacks" into the belief to make it more attitude congruent. The time lag for such causal dynamics to operate is likely micro-seconds and too short to assess longitudinally.

To address the matter, the researcher identified and included in the study an instrumental variable for the positive belief and an instrumental variable for the global attitude. I identify instrumental variables using the letter Z. Recall from Chapter 6 that an

instrumental variable for the M→Y link is one where Z influences M (the **relevance** criterion) but it has no *direct* causal effect on Y (**the exclusion restriction**). To be sure, Z can influence Y but only through M, not independent of M. As well, the instrument should be uncorrelated with the disturbance term for M and the disturbance term for Y (the **independence** criterion, also sometimes called the **exogenous** criterion). This condition would be violated, for example, if there is an unmeasured variable U that influences both Z and M or both Z and Y thereby introducing correlated disturbances between them. I elaborate exceptions to these rules below but for now, I assume these conditions are met in the current example.

Multiple indicators were obtained for some of the variables in the study to create latent variables to adjust for measurement error. All of the measures were multi-item scales that, when items were averaged, had total scores ranging from -3 to +3 with standard deviations near 1.0. The treatment condition was scored 0 = control vs. 1 = intervention. Figure 16.1 shows the relevant influence diagram. The substantive path coefficients of interest are labeled with *p*s, factor loadings are labeled with *L*s, and covariate paths of lesser substantive interest are labeled with *b*s. To avoid clutter, the correlations between the treatment condition and the other exogenous variables are not shown but they are modeled in the Mplus analyses I report later.

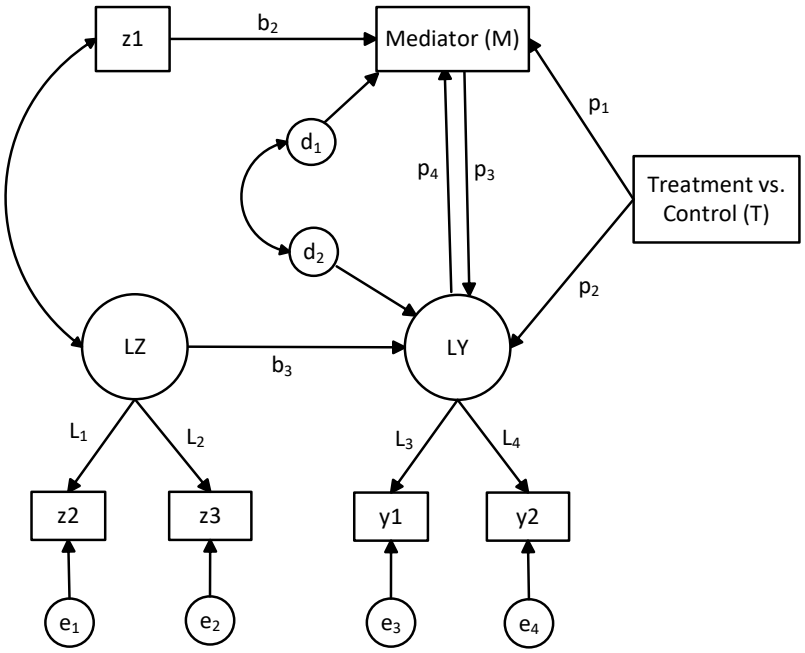


FIGURE 16.1. Reciprocal causality example

Loops in Contemporaneous Reciprocal Causal Models

The key to modeling contemporaneous reciprocal causation using SEM is to frame those effects using the concept of loops. **Loops** are indirect effects composed of two or more direct effects such that the resultant indirect effect ultimately leads back to the originating cause in the indirect relationship. In the current case, the path coefficients p_3 and p_4 in Figure 16.1 satisfy this condition. The direct effect of M on LY in Figure 16.1 is indexed by the value of p_3. However, the resultant change in LY that derives from this direct effect ultimately loops back to M to create additional change in M via p_4. This additional change in M then impacts LY over and above the initial direct effect that M has on Y. The magnitude of this additional effect of M→LY→M→LY is the product of p_4*p_3 yielding a total causal effect of M on Y after one cycle of

$$p_3 + p_4*p_3$$

However, the reciprocal causal dynamic does not stop there. The additional change in LY created by p_4*p_3 in a new pass through the loop impacts M yet again and this new augmented change translates into that much more change in LY, yielding a total effect of M on Y of:

$$p_3 + p_4*p_3 + (p_4*p_3)*p_3$$

Note that the term p_4*p_3 constitutes one **cycle** around the loop if you start at M. The term $(p_4*p_3)*p_3$ constitutes a second cycle around the loop. This cycling process, in theory, continues an endless number of times with each additional cycle yielding an additional term to capture the looping dynamic.

I will notate a loop using the bold faced letter **L**, which should not be confused with the non-boldfaced *L* that I use for a factor loading. We often refer to the "value" of a loop. Suppose **L** equals 0.50. This means that of the initial change created in the target variable M, only half of that initial change will return back to the variable through Y after one cycle, in this case p_4*p_3. The change returning back after two cycles will be half of that half or 0.25 (i.e., $(p_4*p_3)*p_3$) and after three cycles it is half of 0.25 or 0.125. After just a few cycles, the values associated with the more distal looping dynamics become small and are virtually ignorable unless the initial loop value is quite strong. This implies that there is a mathematical limit on the cumulative effect resulting from the initial change that occurs in loop-conceptualized reciprocal dynamics (see Bollen, 1989; Hayduk, 2009).

When both path coefficients in the loop are positive, there is an ever-increasing loop effect that diminishes with each successive cycle until the total effect of one variable on

the other settles down into what seems as functionally a single value. When one of the coefficients in the loop is positive but the other is negative, then increases in the one variable cause the other variable to increase but such increases then cause the first variable to decrease. An example of this phenomenon is the link between increases in law enforcement and decreases in crime. Increased law enforcement causes crime to decrease (a negative coefficient) and the lowered crime rates may then cause law enforcement to become more lax (a positive coefficient). The result will be oscillations in the two variables across time with the oscillations being lesser or greater depending on the magnitude of the paths.

This loop-based conceptualization of reciprocal causality is not without controversy. Some theorists argue that reciprocal causality is properly conceptualized as M and Y being jointly determined by simultaneously influencing one another at a single time point. This orientation is contradicted, however, by the dictum that there always must be a time lag between cause and effect no matter how short that time lag maybe. Truly simultaneous causal effects cannot theoretically exist. The loop-based conceptualization, by contrast, recognizes the need for a time lag between cause and effect in principle but holds that the observational interval of the data may be too wide to capture the relevant causal dynamics. In such cases, conceptualizing the reciprocal causal dynamics via feedback loops of the type described above can yield reasonable summaries of the operative reciprocal dynamics (see Bollen, 2026).

A key assumption in such loop-based modeling is that the direct effect causal coefficients between M→Y and Y→M are stable across cycles (Kaplan et al., 2001) or, more informally, that the core causal coefficients do not change by much from one cycle to the next such that they can be viewed as being "functionally" stable. Stated another way, the values of p_3 and p_4 are thought to remain the same throughout the different cycles. Also, for purposes of model identification, we need at least one instrumental variable for each variable in the reciprocal relationship.

It turns out that the cyclic looping that transpires in loop-based models has implications for other predictors in the model that directly affect either of the two endogenous variables in the reciprocal causal relationship. For example if the dummy coded treatment variable affects M, then when it does so, the M→Y→M cycle is activated which will then further affect M beyond the treatment variable's initial impact on M. This fact needs to be taken into account when estimating the T→M effect, a process I illustrate later when I show the Mplus syntax for the numerical example. I find it useful to think of the final causal estimate as reflecting the relationship for T→M, M→Y, and Y→M as the causal effects after they have reached equilibrium.

Hayduk (2009) argues that there does not, in theory, have to be perpetual looping of the type implied by traditional SEM modeling of contemporaneous reciprocal causality. Perhaps just a single cycle operates in the real world or perhaps just two cycles. Hayduk suggests researchers should specify or speculate an *a priori* number of cycles that can then inform loop-based analytics; that forcing perpetual looping onto the process can sometimes introduce coefficient bias. In practice, if **L** is not large, the assumption of perpetual cycling likely is not problematic because the augmented effects for each cycle eventually become so small as to be virtually ignorable after just a few cycles.

In sum, it is common in SEM modeling of contemporaneous reciprocal causality to assume cyclical or looping effects as the reciprocal causal dynamics play themselves out or have played themselves out on a more micro level. Perhaps the number of cycles that operate are few in number or perhaps they are many but in the final analysis, the mathematics of SEM assumes they are endless. Fortunately, the effects of additional cycles beyond the first few tend to be minor unless the strength of the initial p_3*p_4 value is quite large.

The Role of Disturbances in Contemporaneous Reciprocal Causation

In SEM and in contemporaneous reciprocal causal analysis more generally, we traditionally allow for correlated disturbances between the constituent endogenous variables of the loop (see d_1 and d_2 in Figure 16.1). This is because the disturbance of either of the variables contains part of the disturbance associated with the other endogenous variable's disturbance. A correction is needed and introducing a correlation between the disturbances in the model accomplishes this. Relatedly, Hayduk (2006) proposes a re-specification of the squared R associated with the two endogenous variables in the reciprocal loop using what he calls the **blocked-error R squared**. It indicates the proportion of explained variance in each endogenous variable but excludes the other disturbance term as an influencer. The document on my website for Chapter 16 called *Blocked R Squared* shows the Mplus syntax for calculating this statistic.

Analysis of the Numerical Example

I analyzed the data for the numerical example using the Mplus syntax in Table 16.1.

Table 16.1: Mplus Syntax for Cross-Sectional Reciprocal Causality

```
1.  TITLE: Analysis of cross sectional reciprocal causality
2.  DATA: FILE IS recipmainM.txt ;
3.  VARIABLE:
4.    NAMES ARE id z2 z3 m y1 y2 z1 t ;
5.    USEVARIABLES ARE z2 z3 m y1 y2 z1 t ;
```

```
6.  ANALYSIS: ESTIMATOR=MLR;
7.  MODEL:
8.  Lz ;                    !estimate variance of latent exogenous
9.  Lz BY z2 z3 ;           !indicators for latent Lz
10. Ly BY y1 y2 ;           !indicators for latent Ly
11. y1 ; y2 ; z2 ; z3 ;     !estimate measure errors
12. m ON z1 Ly t ;              !linear equation for M
13. Ly on m Lz t ;              !linear equation for Ly
14. Ly ; m ;                    !estimate disturbance variances
15. Lz z1 t WITH Lz z1 t ;      !correlate the exogenous vars
16. m WITH Ly ;                 !covariance between disturbances
17. MODEL INDIRECT:
18. Ly IND t ;              !estimate effect of t on LY
19. m IND t ;               !estimate effect of t on m
20. Ly IND m ;              !estimate effect of m on Ly
21. m IND Ly ;              !estimate effect of Ly on m
22. Ly IND Lz ;             !estimate effect of Lz on Ly
23. m IND z1 ;              !estimate effect of z1 on m
24. OUTPUT: SAMP STAND(STDYX) MOD(4) RESIDUAL CINTERVAL TECH4 ;
```

All of the syntax should be self-explanatory given the material covered in previous chapters. Note that the specification of reciprocal causality occurs on Lines 12 and 13 where `m` and `Ly` are both predictors and outcomes across the two equations. On Line 15, I tell Mplus to estimate the covariances between the three exogenous variables. Normally, Mplus takes these associations into account by default by conditioning the model on the observed exogenous variables. However, with a latent exogenous variable, Mplus assumes the correlation between the exogenous latent predictor and all other exogenous predictors is zero; it is necessary to override this default to allow these covariances to be estimated as part of the model. If this is done for any one pair of exogenous predictors, then it is best statistically to estimate the covariances for all pairs of exogenous predictors (Muthén, 2015; Muthén, Muthén & Asparouhov, 2016). This approach is implemented on Line 15. For binary exogenous predictors, a warning message will be generated, but often the message can be ignored (see my webpage *Resources* tab for Chapter 11).

Lines 17 through 23 estimate a series of total and indirect effects, each of which targets one of the variables in the reciprocal causal relationship (`m` or `Ly`). This syntax is necessary to take into account the looping cycles implied by the reciprocal causality. If the instrumental variables *per se* are not of theoretical interest but are only a methodological device to allow for estimation of reciprocal causality, then you can exclude Lines 22 and 23. However, it does not hurt to include them as they only affect program output, not model estimation.

In terms of global model fit, the chi square test of perfect model fit in the population was statistically non-significant (chi square = 4.35 with 7 degrees of freedom, $p < 0.74$),

which is consistent with the data being reasonably in accord with the model. The RMSEA was < 0.001 with an upper limit for the 90% confidence interval of 0.063. The p value for close fit was not statistically significant (p < 0.91). The CFI was 1.00 and the standardized RMR was 0.011. For localized fit, there were no theoretically meaningful modification indices greater than 4 and no meaningful residuals between the predicted and observed covariances on a cell-by-cell basis. The model fit seems reasonable.

For the latent variables in the measurement model, the focus is on the standardized output. Here are the relevant portions of it:

```
STDYX Standardization

                                                    Two-Tailed
                    Estimate       S.E.  Est./S.E.    P-Value

 LZ       BY
    Z2                 0.913      0.021     43.382      0.000
    Z3                 0.943      0.017     55.744      0.000
 LY       BY
    Y1                 0.937      0.023     41.240      0.000
    Y2                 0.806      0.029     27.759      0.000

Residual Variances
    Z2                 0.167      0.038      4.357      0.000
    Z3                 0.111      0.032      3.463      0.001
    Y1                 0.122      0.043      2.858      0.004
    Y2                 0.350      0.047      7.472      0.000
```

The standardized factor loadings for `LZ` and `LY` all are reasonable in magnitude (0.913, 0.943, 0.937, 0.806), each with low margins of error. The squared standardized factor loadings can be interpreted as reliability coefficients, with only `Y2` being somewhat on the low side. The standardized residual variances reflect measure unreliability or one minus the reliability estimate for the measure.

Here is the unstandardized output for the core parameters of the structural model:

```
MODEL RESULTS

                                                    Two-Tailed
                    Estimate       S.E.  Est./S.E.    P-Value

 LY       ON
    LZ                 0.682      0.071      9.556      0.000
    M                  0.419      0.105      3.992      0.000
    T                  0.052      0.096      0.539      0.590
```

```
 M        ON
    LY                 0.289      0.085      3.415      0.001
    Z1                 0.393      0.071      5.526      0.000
    T                  0.042      0.118      0.357      0.721
```

Here are the total and indirect effects for M → LY requested on Line 20 of Table 16.1.

```
TOTAL, TOTAL INDIRECT, SPECIFIC INDIRECT, AND DIRECT EFFECTS

                                                    Two-Tailed
                    Estimate       S.E.  Est./S.E.    P-Value

    Effects from M to LY

  Total                0.477      0.141      3.385      0.001
  Total indirect       0.058      0.039      1.464      0.143
```

When documenting the effects of the mediator on the outcome, the total effect of M on Y was 0.477 ±0.28 (Critical Ratio (CR) = 3.38, $p < 0.05$). For every one unit that M increases, the mean of Y is predicted to increase by 0.477 *after taking into account the looping effects*. This effect includes both the direct effect of M on Y plus the loop-based indirect effects via the reciprocal looping between M and Y. The portion of the total effect that is due to the indirect looping is 0.058 ±0.28. The portion that is due to the direct effect of M on Y alone is taken from the output section MODEL RESULTS shown earlier and is 0.419 ±0.21.

As noted earlier, the presence of a reciprocal causal relationship between M and Y in the current example affects the estimate of the total impact of the treatment condition on the mediator. This is because when the treatment has its direct effect on the mediator, this initiates the cycling between M and Y which produces yet more change in M (through Y) over and above the direct effect of T→M. The total effect of the treatment on M taking into account both the direct effect and the indirect looping through Y is:

```
TOTAL, TOTAL INDIRECT, SPECIFIC INDIRECT, AND DIRECT EFFECTS

                                                    Two-Tailed
                    Estimate       S.E.  Est./S.E.    P-Value

Effects from T to M

  Total                0.065      0.119      0.547      0.585
  Total indirect       0.023      0.026      0.871      0.384
```

The effect of the treatment on the mediator is 0.065 ±0.24 (CR = 0.55, *ns*). Because

the treatment condition is a dummy coded dummy variable, the value 0.065 is the estimated mean difference on M for the treatment condition minus the control condition. This effect can be decomposed into the part that is due to the loop-based indirect effect via the reciprocal causation between M and Y and the direct effect of T on M. The portion of the total effect that is due to the looping is 0.023 ±0.05. The portion due to the direct effect of T on M is taken from the earlier `MODEL RESULTS` and is 0.042 ±0.24.

The null hypothesis for the omnibus mediation effect can now be evaluated using the joint significance test for the total effect of T→M and the total effect of M→Y. In the current example, the T→M link is "broken" (statistically non-significant) but the M→Y link is non-zero; the null hypothesis that M is not a mediator of Y cannot be rejected because the joint significance result for mediation was not affirmed.

I do not discuss a full-fledged RET analysis for this example. My intent instead is to give you a sense of issues related to modeling reciprocal causality contemporaneously. As you will see later, some longitudinal modeling includes contemporaneous effects so knowing the ins and outs of such analyses is a useful tool to have.

Limitations of Instrumental Variable Analyses

The use of instrumental variables is one strategy for addressing contemporaneous reciprocal causality. However, despite its strengths, it also suffers from limitations. First, it sometimes is difficult to identify instrumental variables that affect the distal endogenous variable in the loop only through effects on the more immediate endogenous variable. Instrumental variable analysis also can inflate standard errors and reduce precision and statistical power, which is a concern when sample sizes are smaller. Introducing instrumental variables into a model also assumes that the overall model is correctly specified, especially those facets of the model surrounding the instruments. Another limitation is that if the path from the instrument of an endogenous variable to that endogenous variable is weak, it can undermine satisfactory estimation. For discussions of tests for weak instruments, see Chapter 6.

Instrumental variable analysis carries with it the property that the estimated causal effects between the endogenous variables in the reciprocal loop can vary depending on the particular instruments used. If one study uses one instrumental variable and another study uses a different instrumental variable, the results of the causal effect estimates in the two studies can differ. Some theorists find this property bothersome. Instrumental variable analysis also can be undermined by measurement error in the instrument. The numerical example I presented sought to lessen this problem by obtaining multiple indicators of one of the instruments to form a latent variable representation of it. Finally, at a more technical level, instrumental variable based estimates have the desirable

statistical property of consistency but they are not necessarily unbiased; their quality thus can be adversely impacted by the use of small sample sizes.

An important concern when applying loop-based SEM is the possibility of the system mathematically "exploding." With loops, each pass around the loop should contribute a diminishing effect to the component endogenous variables. If the size of the additional effect for a pass instead grows with each cycle, the model can lead to nonsensical results as cyclic growth continues on indefinitely. When this occurs, the effect of a variable on an endogenous variable within a loop is said to be undefined. A diagnostic of this problem is when both standardized coefficients in the loop are > 1.0.

Despite these weaknesses, instrumental variables to explore reciprocal causality in contemporaneous data can be useful. There are strategies other than instrumental variables that can be employed (see Chapter 6) but, in my opinion, most of these newer methods require further validation for use in RETs. As an example, a method sometimes used in place of instrumental variables is to assume path coefficients for the reciprocal causal relationship (p_3 and p_4) are equal in value and are constrained *a priori* to be as such in model estimation. I expand further on using instrumental variables below.

Addressing Confounds in Contemporaneous and Longitudinal Designs

In prior chapters I stressed the importance of addressing confounds to help make valid conclusions about causal coefficients. In longitudinal modeling, methodologists make distinctions between time invariant and time varying confounds. A **time-invariant confound** is when the association between, say, a mediator and an outcome is inflated or deflated over and above M's true causal impact on Y because both M and Y share an "artifactual" common cause that does not change over time. Examples of such confounds might be a history of childhood trauma, exposure to certain types of past parenting practices, and past dietary intake. A **time-varying confound** is one in which the changes in the confound over time simultaneously create changes in M over time and Y over time so as to create bias in causal conclusions about the effects of M on Y. For example, suppose high blood pressure (M) is a presumed cause of overall physical health (Y) for a group of seniors. A person's diet can affect both seniors' blood pressure and their overall physical health. Given this, changes in diet can produce changes in blood pressure and also changes in physical health, thereby creating bias in the assessed link between changes in blood pressure and changes in physical health. This is because a portion of the association between these variables is an artifact of dietary changes over time.

A strategy for dealing with both time invariant and time varying confounds that I emphasized in Chapter 2 was to measure them and statistically control for them during the modeling process. However, even when such confounds are unmeasured, they can in

some cases still be dealt with during statistical modeling by working with disturbance terms. Specifically, if a confound, such as biological sex, is unmeasured but influences both M and Y, then it typically will reside in the disturbance terms of both the cause (M) and the effect (Y). This dual presence results in the disturbance terms for the cause and the effect being correlated. yielding a scenario that appears diagrammatically in Figure 16.2a. If the disturbance terms in Figure 16.2a are left uncorrelated during the modeling process, then their correlation can be inappropriately absorbed into the estimated path coefficient *p*, producing bias in the sample estimate of the population *p*. By simultaneously modeling both the correlated disturbance in conjunction with the causal path *p*, unmeasured confounds are taken into account (but with some exceptions discussed later). To be sure, the unmeasured confounds are not "held constant" but the estimate of *p* is adjusted to account for them.

The problem with this strategy is that the model in Figure 16.2a as presented is under-identified; there are two unknowns (the correlation between the disturbances and the value of *p*) but only one known (the correlation/covariance between M and Y). The model can't be estimated and if you tried to do so in Mplus, you would obtain an error message indicating an under-identified model.

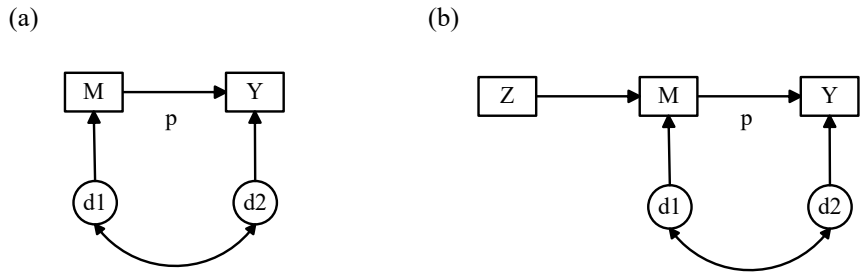


FIGURE 16.2. Dealing with unmeasured confounds

A solution to this problem is to introduce into the model an instrumental variable, Z, directed at the cause in the relationship of interest per Figure 16.2b. To satisfy the conditions of being an instrumental variable in this case, Z should have a non-trivial impact on M, it should not have a direct effect on Y, and it should be uncorrelated with the disturbance terms of the endogenous variables. I created hypothetical data in which the true causal coefficient linking M to Y was 0.50 and where an unmeasured confound, U, had a causal coefficient to M of 0.50 and a causal coefficient to Y of 0.50.[1] When I fit

[1] The variances of M, Y, Z and U were all set to 1.0, which makes the metric analogous to a standardized metric.

a simple bivariate model that linked M to Y and ignored U and the correlation between disturbances, the estimated effect of M on Y was 0.75, which is biased; it should be 0.50. The population data I created were such that there was an instrumental variable, Z, in the data set that fit all the criteria of an instrument, with a causal effect of Z→M of 0.50. When I fit the model in Figure 16.b to the data, the estimated effect of M on Y was 0.50 and the correlation between the disturbances was 0.25. Figure 16.2b adjusted for the unmeasured confound when estimating the target causal coefficient.

Although correlating disturbances can be helpful for dealing with unmeasured confounds, with some exceptions, you probably are better off measuring and statistically controlling for what you think are the primary confounds during your statistical modeling rather than relying on instrumental variables. This is because invoking instrumental variables can inflate standard errors and are not without their shortcomings per my earlier discussion for reciprocal causality. If there are indeed major confounds in your target causal relationship that you are unable to measure and control, then by all means consider instrumental variable strategies to introduce correlated disturbance and address them; but do not invoke the strategy lightly. A bad instrument can wreak havoc on model analysis.

Conditional Instrumental Variables

Sometimes, instrumental variables occur naturally in your RET design and they can be readily exploited accordingly. To recognize them, keep in mind the core properties of an instrument, namely the exclusion restriction, the relevance restriction, and the independence restriction. Parenthetically, the exclusion restriction is often symbolized in the literature using the symbol ⫫ to stand for "is independent of" and the symbol | to stand for "is conditional on" or "holding constant." It is represented mathematically as Y⫫Z|X, which reads "Y is independent of Z holding X constant." The relevance restriction is such that the instrument Z should <u>not</u> be independent of X. It is expressed mathematically as Z⫫̸X.

For the case where multiple variables are part of one's model, there is a special type of instrumental variable that can be invoked when the above properties are not fully met using what are known as **conditional instrumental variables** (CIVs). A CIV is a variable that may not be a suitable instrument when considered unconditionally but it becomes a viable instrument when certain covariates, C, are controlled. Specifically Z is a valid CIV when (1) Z is independent of Y when X *and* C are held constant, (2) Z is not independent of X *when C is held constant*, and (3) Z is independent of the disturbance term that contains the unmeasured confound U when X *and* C are held constant. There also should be no path from Y→X, i.e., no contemporaneous reciprocal causality.

Bollen (2012) describes a model that uses CIVs based on logic developed by Brito and Pearl (2002). Figure 16.3 presents the influence diagram for the model:

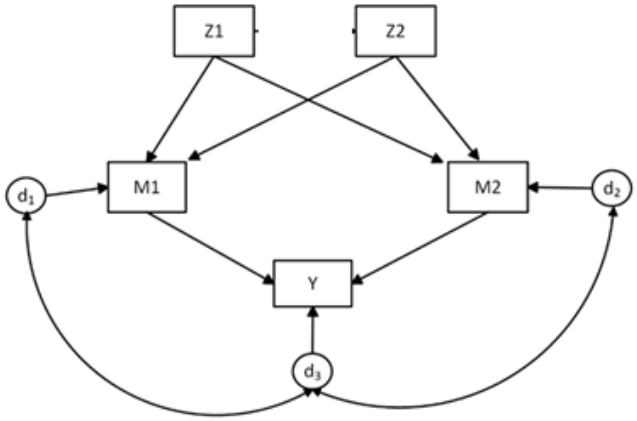


FIGURE 16.3. Example of conditional instrumental variables

The SEM-based equations for the endogenous variables that describe this model are

$$M1 = a_1 + p_1\, Z1 + p_2\, Z2 + d_1 \quad [16.1]$$

$$M2 = a_2 + p_3\, Z1 + p_4\, Z2 + d_2 \quad [16.2]$$

$$Y = a_3 + p_4\, M1 + p_5\, M2 + d_3 \quad [16.3]$$

Z1 serves as a CIV for the M1→Y relationship if we assume the model is true because (a) it impacts M1 *and* (b) when we hold M2 constant in Equation 16.3, Z1 functionally becomes independent of Y. If M2 is not held constant, then Z1 would not be a viable instrument for M1→Y. By the same token, Z2 also is a CIV for M1→Y as are both Z1 and Z2 for M2→Y. The correlated disturbances in the model should be identified because of the presence of the Z1 and Z2 CIVs.

Table 16.2 presents Mplus syntax for fitting the model in Figure 16.3. (The measures all had standard deviations near 1.0, to give you a sense of their metrics).

Table 16.2: Mplus Syntax for Conditional Instrumental Variable

```
1.  TITLE: Analysis of conditional instrumental variables
2.  DATA: FILE IS civ.txt ;
3.  VARIABLE:
4.    NAMES ARE m1 m2 y z1 z2 ;
5.    USEVARIABLES ARE m1 m2 y z1 z2 ;
```

```
6.  ANALYSIS: ESTIMATOR=MLR;
7.  MODEL:
8.  !define equations
9.  m1 on z1 z2 ;
10. m2 on z1 z2 ;
11. y on m1 m2 ;
12. !estimate disturbance variances
13. m1 ; m2 ;  y ;
14. !estimate covariances
15. m1 with y ; m2 with y
16. OUTPUT: SAMP STAND(STDYX) MOD(4) RESIDUAL CINTERVAL TECH4 ;
```

The correlated disturbances are requested on Line 15. The data set is available for download on the *Resource* tab of my website.

When I fit the model to the data, the model yielded reasonable global fit (chi square test = 1.80 with df = 1, $p < 0.18$; RMSEA < 0.037, 90%CI = 0.00 to 0.12; p value for close fit <0.46; CFI = 1.00, SRMR = 0.001). There were no modification indices greater than 4 and all of the standardized residuals were reasonable in magnitude.

I first check the coefficients from Z1 and Z2 to M1 and M2 to ensure they are not weak instruments. Here are the relevant statistics:

```
MODEL RESULTS

                                                    Two-Tailed
                  Estimate       S.E.  Est./S.E.    P-Value

M1       ON
    Z1               0.470      0.037     12.541      0.000
    Z2               0.210      0.042      4.987      0.000

 M2       ON
    Z1               0.147      0.038      3.831      0.000
    Z2               0.540      0.043     12.560      0.000
```

A common standard for evaluating instrument strength is that the *squared* critical ratio for the instrument to the target variable should be greater than 10, which was the case here. For example, the squared critical ratio for Z1 to M1 is 12.541^2f and for Z2 to M1 it is 4.987^2. Also, the p values all were statistically significant and the coefficients seemed reasonable in magnitude given the standard deviations of the measures.

Another important property is that the instrumental variable(s) be relatively uncorrelated with the disturbance term of the relevant endogenous variable, which is the exogeneity restriction. For example, Z1 should be uncorrelated with d2 in Figure 16.3 as should Z2. Unfortunately, this is untestable unless you have multiple instruments for the

target endogenous variable. In the current model there are two instruments (Z1 and Z2) for M1 and two instruments (Z1 and Z2) for M2 which means I can use the **Sargan test** to gain perspectives on the exogeneity restriction. The Sargan test focuses on the null hypothesis that *all* instrumental variables relevant to the M1→Y link are uncorrelated with the Y disturbance term versus the (undesired) alternative hypothesis that at least one of the instrumental variables correlates with the disturbance. Significant p values for the Sargan test tell us that at least one of the instrumental variables correlates with the equation error, but it does not identify which one.

The Sargan statistic approximates a central chi-square distribution with degrees of freedom equal to the number of instruments minus one. In general, we want the Sargan test to yield a p value that is statistically non-significant. I applied the test using two stage least squares in the R package *MIIVsem* which is available on my website. *MIIVsem* interfaces with the R package lavaan and incorporates lavaan syntax. Here is the syntax I used within *MIIVsem* (see the *Syntax* tab on my webpage for how to program in lavaan):

```
y ~ m1 + m2
m1 ~ z1+z2
m2 ~ z1+z2
m1 ~~ y
m2 ~~ y
```

Here is the output that contains the Sargan test:

```
STRUCTURAL COEFFICIENTS:

                   Estimate  Std.Err  z-value  P(>|z|)   Sargan   df   P(Chi)
  m1 ~
    z1                0.470    0.037   12.541    0.000    1.797    1    0.180
    z2                0.210    0.042    4.987    0.000
  m2 ~
    z1                0.147    0.038    3.831    0.000    1.797    1    0.180
    z2                0.540    0.043   12.560    0.000
```

Neither of the Sargan tests were statistically significant, which is reassuring.[2]

Given support for the instruments, here are the core regression coefficients and covariances in the model as reported by Mplus for the main model analysis:

[2] Parenthetically, the Sargan test has some shortcomings so it must be interpreted tentatively (see Jin, 2022).

```
MODEL RESULTS

                                                    Two-Tailed
                  Estimate       S.E.  Est./S.E.    P-Value

 Y        ON
    M1              0.333      0.097      3.445      0.001
    M2              0.467      0.098      4.786      0.000
 M1       WITH
    Y               0.358      0.104      3.443      0.001

 M2       WITH
    Y               0.236      0.108      2.182      0.029
```

Both mediators were statistically significant predictors of the outcome and both of the covariances for the correlated disturbances were statistically significant. The values of the path coefficients for M1 and M2 to Y are adjusted for unmeasured confounds by virtue of including the correlated disturbances. Here are the two relevant covariances expressed as correlations to improve interpretability, obtained from the standardized output section:

```
                                                    Two-Tailed
                  Estimate       S.E.  Est./S.E.    P-Value

 M1       WITH
    Y               0.366      0.097      3.776      0.000

 M2       WITH
    Y               0.236      0.107      2.199      0.028
```

It is interesting to compare the coefficients to what happens if I ignore the dynamics of the instrument. Here are the results where I omitted the correlated disturbances:

```
MODEL RESULTS

                                                    Two-Tailed
                  Estimate       S.E.  Est./S.E.    P-Value

 Y        ON
    M1              0.587      0.034     17.459      0.000
    M2              0.606      0.033     18.284      0.000
```

The coefficients are biased upward because they include the unmeasured confounds.

People often bemoan the difficulty of finding viable instrumental variables to include in one's model but sometimes instruments naturally occur as part of the RET design. Figure 16.4 presents an influence diagram of a single mediator longitudinal RET

in which the time of assessment is denoted with a subscripted *t* followed by the wave number; for example, t_1 is baseline, t_2 is the wave 1 posttest, t_3 is the wave 2 follow-up and t_4 is wave 3 when the outcome is assessed. It turns out this particular RET design inherently contains instrumental variables that permit estimation of the causal paths and correlated disturbances simultaneously if doing so can be theoretically justified. I have inserted dashed curved arrows in Figure 16.4 between disturbance terms where I could, if desired, introduce correlated disturbances into the model without compromising the identification status of the model. The baseline $M1_{t1}$ is an instrument for the effect of M_{t2} on M_{t3} (as is T) and M_{t2} is an instrument for the effect of M_{t3} on Y. Keep in mind that using lagged variables as instruments requires the usual assumptions of instrumentation, namely that the lagged instrument affects the target endogenous predictor (relevance), that it only affects the outcome through the endogenous predictor (exclusion) and that it is uncorrelated with the endogenous error terms (independence). If these conditions fail, then the lagged instruments fail and can introduce bias if treated as instruments by correlating the disturbances. Wang and Bellmare (2020) are skeptical about the viability of the independence assumption when using lagged outcomes as instruments and recommend against using them in this way. For the model in Figure 16.4, this raises questions about including the identified lagged predictors with their correlated disturbances.[3] However, this does not preclude using the lagged variable by and of itself without the corresponding correlated disturbance.[4] I revisit this issue below.

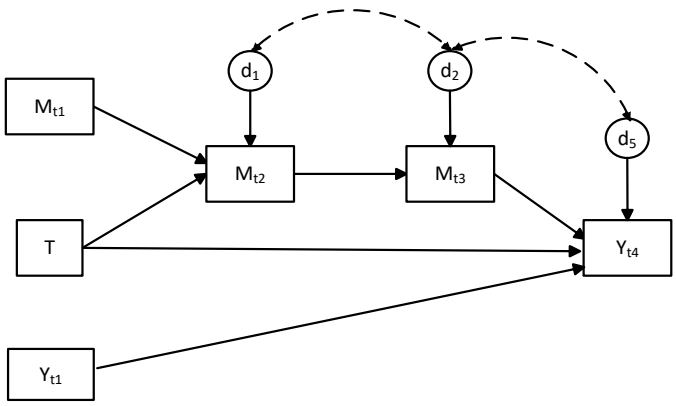


FIGURE 16.4. RET design with instrumental variables

[3] Allison (2018) show that in SEM, for the lagged variable to take on instrumental properties, you must include both the path between the target endogenous variables *and* the correlation between their disturbances.
[4] Allison (2018) show that in SEM, for the lagged variable to take on instrumental properties, you must include both the path between the target endogenous variables *and* the correlation between their disturbances.

As a final note, do not fall prey to the critic who waves the general wand of unmeasured confounders to dismiss RET results. Such critics, in fairness, should be able to articulate the specific confounds they think undermine your conclusions and build a substantively logical case for non-trivial effects of those confounds. Having said that, I also think you should always approach your causal conclusions with humility because there could, in fact, be unmeasured confounds that *are* leading you astray. For additional perspectives on instrumental variables to deal with confounds, see Angrist, Imbens & Rubin (1996), Bollen (2015), Henckel, Buttenschoen & Maathuis, 2024, Imbens (2014) and van der Zander, Textor & Liskiewicz (2015).

Autoregression in Longitudinal Designs

Another important concept in longitudinal modeling is that of autocorrelation and its counterpart, autoregression. **Autocorrelation** is the degree of correlation between the same variable across time. It reflects how the lagged version of a variable is related to the current version of that variable. **Autoregression** regresses the current values of a variable onto a lagged version of that variable, usually in the form of a linear model, as follows:

$$Y_t = a + b\ Y_{t-1} + d \qquad [16.4]$$

where *t* refers to time *t* and *t-1* is an *a priori* specified time before *t*. The letter *a* is the intercept, *b* is the slope or path coefficient, and *d* is the disturbance term in the linear equation, expressed here using sample notation. As an example, I might measure someone's weight, Y, on 4 different occasions at 3 month intervals. The different measures can be referred to as Y_{t1}, Y_{t2}, Y_{t3} and Y_{t4}. For any given measure of weight taken at time *t*, I can refer to the previous measure of weight at *t-1*. Note that this lagged value does not exist for Y_{t1} because I did not obtain any measure of Y prior to *t1*.

Equation 16.4 is a way of saying mathematically that peoples' weight at a given point in time, *t*, can be predicted from their weight at the time measured one lag before it, i.e., from *t-1*. Autocorrelation is the sign and magnitude of the correlation that results from applying Equation 16.4. The **autoregressive** coefficient is the slope parameter, *b*. It tells us for every one unit that Y_{t-1} increases, how much the mean of Y at time *t* is predicted to increase or decrease.

Autoregressive paths are included in longitudinal models for a variety of reasons. One reason is that the status on a variable at time *t-1* can often serve as a proxy that when statistically held constant blocks the pathways from more distal across-individual confounds that impact the variable at time *t*. The lagged Y renders those distal confounds harmless as contaminants of other determinants of the time *t* variable (see Chapter 2 for

elaboration). This has been our major use of the baseline outcome in prior chapters, namely as an economical way of confound control. We are not particularly interested in the autoregressive coefficient per se nor in imposing interpretations onto it. It is merely a way of controlling confounds or of increasing the statistical power of statistical tests.

A second reason we include lagged versions of a variable in our models is that the autoregressive coefficient and/or correlation linking the time *t* variable to that same variable at time *t-1* is thought by many to reflect across-individual construct stability, a phenomenon that may be of substantive interest and that we want to make statements about. As I show below, this attribution is more complicated than that.

Third, a lagged outcome variable might be included in a model because the variable's standing at time *t-1* can impact the standing of a person at time *t* on the outcome variable in a causal sense independent of its role in blocking confounds of across-individual stability. Stated a bit more formally, peoples' departures at time *t-1* from their general overall mean score on a variable over time can carry over in a causal sense to a later assessment, a dynamic often called an **inertia or carry-over effect**. A person's mood at time *t-1* may carry over to shape the person's mood at time *t*.

Finally, a lagged variable may serve as a baseline from which the investigator wants to document future change and thereby serves as a benchmark for documenting change.

Models with autoregression sometimes extend beyond one lag. Consider the following model that measures Y at four time points at three month intervals:

$$Y_t = a + b_1 Y_{t-1} + b_2 Y_{t-2} + d \qquad [16.5]$$

This model implies that peoples' Y scores at time *t* are a function of their Y scores three months prior to time *t* (namely Y_{t-1}) and, independent of this, their Y scores at *t-2* or Y_{t-2}. It is as if there is a delayed effect relative to the three month time interval used in the study such that it takes longer than 3 months for earlier differences in Y at *t-2* to translate into Y_t. These delayed effects do not fully manifest until, say, 6 months have passed, not just three months. When only a single lag is included as a predictor in the model, it is referred to as a **first order autoregressive model**. When two successive lagged variables are included as predictors in the model, it is referred to as a **second order autoregressive model**. A **third order autoregressive model** would be where a third lag Y_{t-3} is included in the model as a separate predictor for Y at time *t* independent of Y at times *t-1* and *t-2*. And so on. Note that in an empirical study of autoregressive dynamics with four time points, a second order autoregressive model is only defined for Y_{t4} and Y_{t3} because we do not have information on the two prior lags for variables Y_{t1} and Y_{t2} in the data.

The Simplex Structure

A common longitudinal structure for autoregression is known as the **simplex model**. The idea is that autoregressive coefficients between successive lags are equal in value, per Figure 16.5, which shows coefficients for lagged variables using a standardized metric.

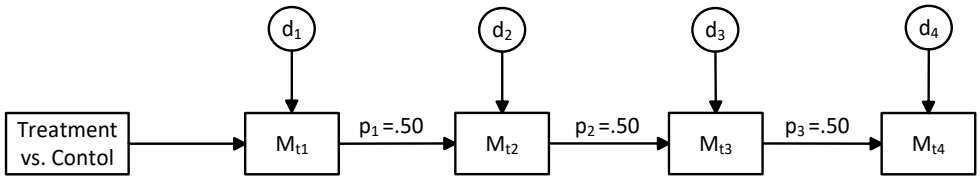


FIGURE 16.5. RET with mediators implying a simplex correlation structure

If I apply the decomposition process described in Chater 7, it can be shown using algebra that the correlation between M_{t1} and $M_{t2} = p_1 = 0.50$, the correlation between M_{t1} and $M_{t3} = p_1 * p_2 = (0.50)(0.50) = 0.25$, and the correlation between M_{t1} and $M_{t4} = p_1 * p_2 * p_3 = (0.50)(0.50(0.50) = 0.125$. Note that the more distal the variable is in time from a future version of that variable, the lower the correlation between them will be, which makes sense for a wide range of phenomena. Given this, many theorists impose equality constraints on first order autoregressive coefficients that are equally spaced or nearly so, per Figure 16.5. Of course, the simplex pattern can be disrupted if there are unmeasured time-varying confounds that create unequal correlated disturbances, per Figure 16.6.

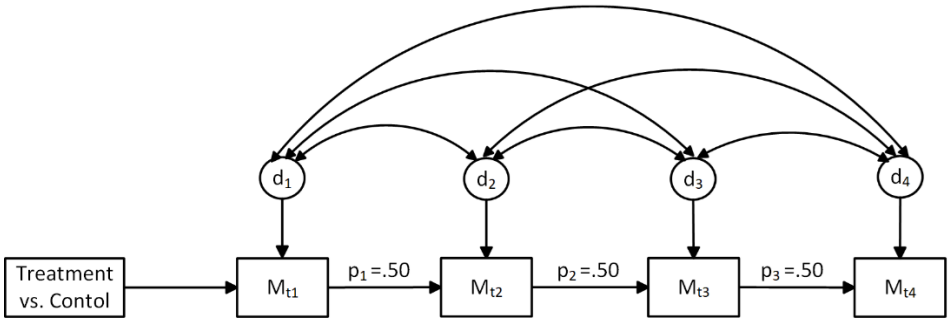


FIGURE 16.6. RET with mediators with correlated disturbances

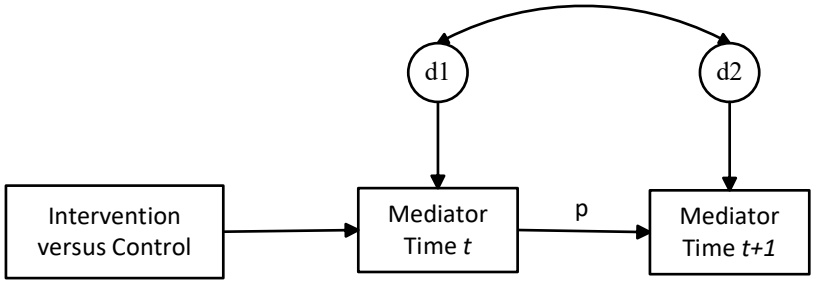
Autocorrelation, Autoregression, and Stability

It is common for researchers to interpret autoregressive coefficients as stability coefficients because they supposedly reflect how stable a construct is over time. The higher the autocorrelation, the more stable the construct is thought to be, everything else being equal. Although there is some truth to this interpretation, such characterizations oversimplify matters. For example, suppose Y is measured on a 0 to 10 metric and in response to an intervention everyone shifts their Y score downward by two units. The correlation between Y and the prior (baseline) measure of Y will be 1.0 but Y is not stable; everyone has changed by 2.0 units. In addition to the correlation between Y_t and Y_{t-1}, one also must examine the mean structure of the two measures to make attributions of stability. In addition, some individuals can show shifts in their Y scores between Y_t and Y_{t-1} but the rank order of individuals on Y at a given time point remains unchanged at future times. This yields a Spearman rank order correlation of 1.0 and implies the rank ordering of individuals is "stable" despite the fact that changes in the raw scores per se across time may have occurred. Finally, in many measures we use, there are degrees of random noise in any given person's score that reflect measurement unreliability. This means a person whose true score is stable across time may exhibit different observed scores or that a person's whose true scores differ across time show the same observed scores because of random error affecting the observed scores.

Autoregression and Causal Inference

In longitudinal modeling it is common to include autoregressive links because they are thought to reflect a formal causal connection between Y at time t and Y at time $t+1$. However, in many cases, the association between Y_t and Y_{t+1} reflects something other than the causal effect. Usually, the association also reflects the influence of a common cause from a stable, time-invariant variable on both Y_t and Y_{t+1} . For example, if Y_t is influenced by biological sex and if Y_{t+1} also is influenced by biological sex, this dynamic will contribute to a correlation between Y_t and Y_{t+1}. This confound, in turn, inflates (or deflates) the association between Yt and Y_{t+1} so that the magnitude of the association does not just reflect the casual link between them; it also reflects the common cause confound of biological sex. If I ignore the common influence of this stable, time-invariant variable on Y_t and Y_{t+1} and just regress Y_{t+1} onto Y_t in the spirit of autoregression, then the resulting path coefficient will be a biased estimate of the true causal coefficient linking Y_t to Y_{t+1}. The role of stable, time-invariant confounds has been increasingly recognized as a potential flaw in traditional cross-lagged models that have been used for many years to infer lagged effects. I discuss cross-lagged models in depth below.

Confound control for autoregressive effects often can be achieved by identifying the most important potential confounds, measuring them, and statistically controlling for them. The strategic introduction of correlated disturbances also is an option. The basic form the latter takes is as follows:



The goal is to obtain an unbiased estimate of *p*. Consider first the case where there is an unmeasured potential confounder that is time-invariant, say, biological sex. Suppose that sex influences M_t but it does not impact M_{t+1} over and above the effect of M_t on M_{t+1}. In this case, the correlated disturbances are not needed because the effect of biological sex on M_{t+1} is completely mediated by M_t. It is not a confounder in the sense that once I control for the effect of M_t on M_{t+1}, biological sex is controlled for. Suppose, however, that M_t does not fully mediate the effect of sex on M_{t+1}. Sex now resides in the disturbance terms for both M_t and M_{t+1}. I need to add the correlation between the disturbance terms using SEM to account for this and to adjust for the bias that would otherwise manifest itself in *p* by virtue of the shard variance being absorbed into *p*.

The above strategy works reasonably well for time-invariant confounders. However, it can encounter difficulties with time-varying confounders, especially with three or more waves. Suppose the unmeasured confounder, U, is experienced stress and that M is mood. Suppose further that stress from time *t* dissipates to some extent at time *t+1*, hence it is time-varying. I might conjecture that mood at time *t* impacts mood at time *t+1* in the form of a carry-over effect. The disturbances of both M_t and M_{t+1} will both stress (U) in them but the stress will be lower in the disturbance of M_{t+1} relative to M_t. I expect the disturbances to be correlated because stress is in each of them. However, the underlying causal dynamics can be more complex that what a simple correlated disturbance parameter can account for. U at time *t* may impact U at time *t+1* and M at time *t* may impact U at time *t+1*. The unmodeled and complex dynamics can breach the exclusion restriction assumption for instrumental variables, the result being bias in the estimation of *p*. To free estimation from potential bias due to time-varying confounds, one must either include measures of the time-varying confounds in the model as

covariates to control for them or assume their impact is minimal.

When you include correlated disturbances in conjunction with autocorrelation, the autoregressive path coefficient is interpreted as the predicted change in Y*t+1* given a one-unit change in Y*t* net any shared background variance between the two time points.

The Controversy of Lagged Dependent Variables

Directly relevant to the above is the rather large literature on the inclusion of lagged dependent variables in panel models. Objections to the practice in one form or another have been offered by Achen (2000), Bruderl and Ludwig (2014), Dafoe (2018), Foster (2010), Keele and Kelly (2006), Leszczensky and Wolbring (2022), Morgan and Winship (2014), Mouw (2006), and Vaisey and Miles (2017), among others. Some statements arguing against the use of lagged outcomes are rather strong, almost overly so in my opinion. The argument is that including lagged outcomes can artificially introduce bias into other path coefficients and that it is better to avoid this bias by not the including lagged outcomes. Note that these objections are not necessarily to using lagged variables as instruments. Rather, they are objections to using lagged outcomes at all.

As discussed in previous chapters, the arguments *in favor* of including lagged variables are that (*a*) they can serve as proxies to control for more distal across-individual unmeasured confounds that impact the target variable at time *t* (*b*) the lagged coefficient provides perspectives on construct stability, which may be of substantive interest, (c) the autoregressive coefficient may capture inertia effects of the lagged variable on the target variable at time *t*, resulting in bias if the lagged variable is omitted and (d) the lagged variable may serve as an interesting baseline from which to document change.

The arguments against inclusion of lagged outcomes primarily (but do not exclusively) center on their behavior in the presence of certain forms of unmeasured confounds. The underlying dynamics have been characterized in different ways for different modeling scenarios but the essence of many of the arguments are captured in Figure 16.8 (see Andersen & Mayerl, 2023, for elaboration). The variable X in Figure 16.8 is the causal determinant of interest. Y_{t2} is the outcome variable of interest and Y_{t1} is the lagged version of it. The coefficient p_1 is an autoregressive coefficient and p_2 is the causal coefficient we seek to estimate. There are two unmeasured confounding variables U1 and U2 which I put in dashed boxes to emphasize that they are unmeasured and hence not statistically controlled. I omit disturbance terms to avoid clutter but they are implied.

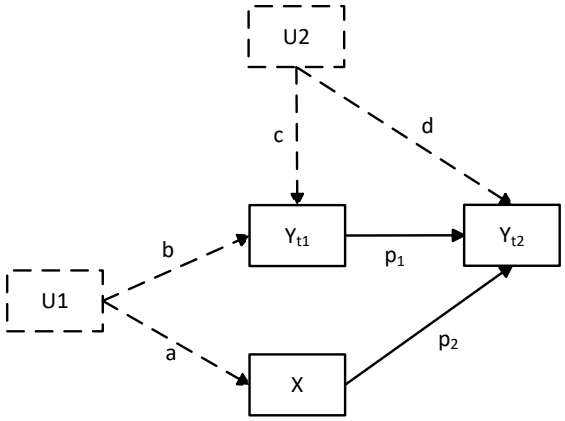


FIGURE 16.8. Lagged outcomes and unmeasured outcomes

The variable U2 is a time-invariant confound for the $Y_{t1} \rightarrow Y_{t2}$ effect. If U2 impacts both Y_{t1} and Y_{t2}, we normally would expect that by statistically controlling for Y_{t1}, we block the effect of U2 on Y_{t2}. However, in this particular case, U2 has an independent effect on Y_{t2} over and above Y_{t1} (path *d*), a property that may or may not apply in practice. U1 is an unmeasured confound for the X to Y_{t1} relationship.[5] Note that X could be a mediator or it could be a randomly assigned treatment condition. If it is the latter, then we would expect path *a* to be zero or functionally zero due to random assignment, thereby rendering the confound irrelevant assuming random assignment is properly implemented.

It can be shown mathematically that if both U1 and U2 are present with non-trivial values for paths *a* through *d* in the manner shown in Figure 16.8, then statistically controlling for the lagged Y_{t1} when regressing Y_{t2} onto X will introduce bias into the estimate of p_2. For mathematical details, see Andersen and Mayerl (2023). Note that this bias would not occur if we identified and measured U1 and U2 and statistically controlled for them. Nor would it matter if any of these paths are weak and inconsequential. Dafoe (2018, p. 139) notes that the presence of both U1 and U2 in the model creates a collider dynamic (see Chapter 2) but that it otherwise is safe to use the lagged outcome (a) if U1 is not operative or if it is rendered moot via statistical controls or (b) if U2 is not operative or rendered moot. This underscores the importance of trying to identify U1 and U2 *a priori* so they can be measured and statistically controlled.

Keele and Kelly (2006) argue that when a lagged outcome is, in fact, causally linked to the outcome, then excluding it from a model out of fear of collider bias can

[5] Whether the model includes a causal link between X and Y_{t1} is immaterial to the current arguments, so I omit it here.

create omitted variable bias in its own right. Including the lagged outcome in such contexts often produces estimates with less bias compared to model forms that include the lagged variable, with the exception being the presence of rather large effects of U1 and U2. Wilkins (2017) shows that the inclusion of a lag 2 outcome as a predictor in conjunction with a lagged X (i.e., regress Y_{t2} onto $Y_{t-1} + Y_{t-2} + X_t + X_{t-1}$) will often introduce controls that address the offending collider confounds. This strategy is a special case of what is known as **autoregressive distributed lag** (ADL) modeling. Leszczensky and Wolbring (2022) also emphasize the benefits of using both contemporaneous (X_t) and single lag (X_{t-1}) predictors of Y_t to adjust for bias when including Y_{t-1} as a predictor (see also Murayama & Gfrörer, 2024).

Allison (2022) considered the case of coefficient bias with lagged outcomes for the case of certain forms of misspecified models, namely the case where (a) the true model has no autoregressive effect of a prior Y on Y and where Y at time *t* depends only on X at the same time point but (b) the researcher estimates the coefficient, β, for the regression of Y only on X at time *t*-1. It turns out the expected value of the coefficient for X_{t-1} in this case is exactly -.5β. In other words, a positive contemporaneous effect of X on Y translates into an artifactual negative lagged effect in the misspecified model. For statistical details, see Allison (2022). The implication is that if you are analyzing models with lagged relationships and you observe counter-intuitive sign reversals for the coefficient for the lagged predictor(s), be wary of specification error. Note also that if the true model is, in fact, one in which X_{t-1} impacts Y_t but there is no contemporaneous effect of X_t on Y_t, then if you regress Y_t onto both X_{t-1} and X_t, the estimated contemporaneous effect of X_t on Y_t should be small and not statistically significant. Allison suggests modeling strategies to help protect against lagged sign reversals and I address these below. In the final analysis, it ultimately is important that you can make a good case substantively that your model does indeed capture the correct lag structure.

In sum, there is controversy surrounding the use of lagged outcome variables when working with panel data. Reasonable arguments have been offered for both the inclusion and exclusion of them. The bottom line is that difficulties with lagged outcomes often derive from unmeasured confounds or misspecified models but there are remedies that usually permit one to accommodate lagged outcomes either through measuring and statistically controlling for confounds or by the judicious introduction of correlated disturbances. My own view is that in RETs, using outcome baseline values to gain statistical power in T→M or T→Y assessments is reasonable and is protected by random assignment. The baseline outcome as a covariate also plays a critical role in the control of distal confounds for M→Y analyses (per Chapter 2) and should be considered accordingly. If my RET expands into a multi-wave panel model where autoregressive

effects beyond the baseline may be required, I adopt what I call the AMM approach to RET design: *articulate, measure, and model*; Articulate the most important confounders that need to be controlled, make a good faith effort to validly measure them, and use state of the art modeling to then statistically control them.

Measurement Invariance Across Time

When we model path/autoregressive coefficients or changes in a construct over time, we make assumptions that the measurement properties of variables remain unchanged; that any changes in the observed scores reflect changes in the underlying constructs the variables are thought to measure. If I observe a mean change in the observed scores for a rating scale across two time points, it could be (a) that the mean of the measured construct has changed and/or (b) that the way people interpret the rating scale has changed but the mean has not. In the latter case, what we think is true change in the construct actually is an artifact of changes in the measurement properties of the construct indicators. If we have multiple interchangeable indicators of a construct, we can formally evaluate the invariance of measurement properties across time. I discussed this matter in Chapter 3 and describe ways of implementing tests of measurement invariance over time in the document *Measurement Invariance* on my website for Chapter 3

Consider the autoregressive model in Figure 16.7. Longitudinal measurement invariance implies that parallel factor loadings across time are equal, that is $L1_{t1} = L1_{t2} = L1_{t3}$; $L2_{t1} = L2_{t2} = L2_{t3}$; $L3_{t1} = L3_{t2} = L3_{t3}$. In addition, if you compare means across time, the measurement intercepts linking the respective LY to its measurement intercept should be equal across time, that is a_{t1} for $Y1_{t1} = a_{t2}$ for $Y1_{t2} = a_{t3}$ for $Y1_{t3}$; a_{t1} for $Y2_{t1} = a_{t2}$ for $Y2_{t2} = a_{t3}$ for $Y2_{t3}$; a_{t1} for $Y3_{t1} = a_{t2}$ for $Y3_{t2} = a_{t3}$ for $Y3_{t3}$. I discuss how to perform these tests in the document *Measurement Invariance* on my website.

In general, we want to make the strongest statements we can about the compatibility of the measurement properties across time. When possible and when it makes sense to do so, you should evaluate across-time measurement invariance in longitudinal designs.

Some researchers assert the importance of measurement invariance by *a priori* imposing equality constraints on parallel factor loadings across time. This practice is controversial. Strict equality of factor loadings or measurement intercepts across groups or time is a very stringent condition that rarely holds in the real-world. Blithely forcing invariance onto a model does not make non-invariance go away. Instead it usually redistributes the non-invariance to other model parameters that then can create bias in them. It also can lead to poor global model fit that results in the rejection of a potentially viable structural model.

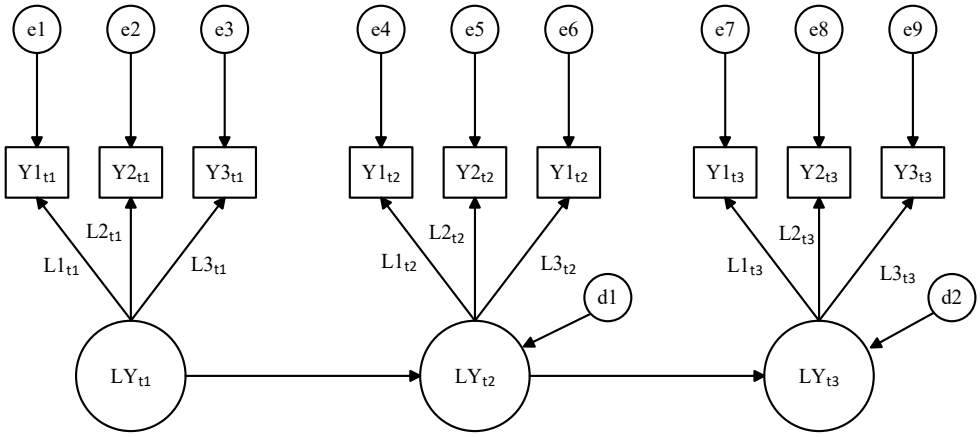


FIGURE 16.7. Latent variable autoregression model

In my opinion, rather than trying to force measurement non-invariance into submission via *a priori* specified equality constraints, it is best instead to explore the matter in depth using the methods described in the document on the *Resources* tab of my webpage for Chapter 3 (see also the document on growth curve modeling called *Additional Applications of Latent Growth Curve Modeling* on my webpage for the current chapter). Research on partial invariance suggests that (a) if the reference indicator is reasonably measurement invariant and (b) if one or more of the indicators other than the reference indicator is reasonably measurement invariant, then ignoring small or modest levels of measurement invariance for the other indicators often will not matter all that much and is more reflective of real world dynamics; the extent to which it does matter can be explored using the approaches described in the document on the *Resources* tab of my webpage for Chapter 3. I think a heathy attitude towards group or time differences in key measurement parameters is to view such differences not necessarily as artifacts but rather as phenomena that require recognition, explanation, and appropriate modeling. Pretending it does not exist by *a priori* imposing strict measurement invariance on our models is probably not the best way to go.

Between-Person and Within-Person Analyses

Most of the analyses I have considered in this book analyze causal relationships from a **between-person** (also called **across-person**) perspective. Between-person analyses seek to identify variables that create differences between individuals. Such explanatory variables make some people higher or lower on an outcome than other people. Suppose I want to determine if there is a link between mood positivity (a mediator) and self-esteem

(an outcome) in a between-person analysis. For each of 100 individuals, I measure their mood positivity and, just afterwards, their self-esteem. I then regress self-esteem onto mood positivity for the 100 individuals. If the positivity mediator (M) has a regression coefficient of, say, 2.2 associated with it, this means that people who score one unit higher on M than other people will, on average, score 2.2 units higher on Y. For a treatment dummy variable (T, scored 0 = control, 1 = intervention) if the regression coefficient for it when predicting Y is 1.8, this means that people who score one unit higher than other people on T will score, on average, 1.8 units higher on Y. If I have controlled for relevant confounds, I might make causal inferences about the variables based on such coefficients, albeit with the usual qualifications.

There is, however, a second framework for making causal inferences and it uses a **within-person** perspective. Within-person analyses identify variables that create differences within an individual across multiple time points or from one occasion to the next occasion. Suppose I measure the positivity of people's mood on 21 consecutive days. I might observe variability in mood positivity for each person across the 21 days and also in self-esteem. I can regress the 21 self-esteem ratings for a person onto the corresponding 21 mood ratings and identify a regression coefficient to isolate the *within individual* effects of changes in mood on self-esteem. Suppose, as in the between-person analysis, I find the typical within person coefficient is 2.2; for every one unit that a person's mood changes across the 21 occasions, the person's self-esteem is predicted to change by 2.2 units. In this case, the coefficient replicates the between-person result.

Note, however, there are properties of the within-person analysis that some researchers find more satisfactory than the between-person analysis. For example, in the within-person analysis, we control for all between-person confounds because the analysis is focused on what is occurring *within* individuals across time. When I assess the link between mood positivity and self-esteem within people, variables like race, SES, and maternal education typically do not change from one occasion to the next. They are time-invariant. As such, they are "held constant" across the time points because they take on the same value at each assessment for a given individual. The within-person coefficient is free of all time-invariant confounds.

In panel models of the type I consider in this chapter, the general notions of between-person and within-person effects can be further understood in a rough sense with reference to Table 16.3. Suppose in an RET I measure both an outcome, Y, and a mediator, M, at an immediate posttest and again at 3, 6, 9 and 12 month follow-up intervals (I address the issue of including or excluding baseline measures later in the chapter). I seek to gain perspectives on whether there is a causal link between M and Y. The left portion of Table 16.3 provides the data for the Y scores for each person and the

right portion provides the data for the M scores. Both M and Y have a metric ranging from 0 to 10. The columns labeled "Mean" are the mean values of Y (or M) for each person across the five measurement occasions, labeled T1 through T5.

The columns labeled 'Mean' represent the overall tendencies of each individual to score higher or lower on Y (and M) across time. One way of indexing the causal link between M and Y on a between-person level is to regress the N=200 mean scores of Y (i.e., the Mean column for Y) onto the Mean column for M. The resulting regression/path coefficient is an estimate of the between-person causal coefficient. It, of course, is subject to contamination by across-person confounds, like ethnicity and SES.

Table 16.3: Hypothetical Scores for M and Y

	Scores on Y						Scores on M					
Person	T1	T2	T3	T4	T5	Mean	T1	T2	T3	T4	T5	Mean
1	3	4	2	6	5	4.00	4	5	3	5	3	4.00
2	3	4	1	2	0	2.00	3	4	1	1	1	2.00
3	6	8	4	6	6	6.00	5	7	3	5	5	5.00
.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.
200	7	6	7	4	1	5.00	7	7	7	5	4	6.00

There is a different approach I can use, however. One way of thinking about the two columns for the individual-based Y and M means is that each score reflects the effects of the more stable backgrounds for Y and M that individuals bring to the study. For the Y scores, suppose the grand mean of the column of Y Means equals 5.0. This value represents the overall "typical" Y score for all individuals in the study. The first person has a "background" mean of 4.0. The effect of his or her background seems to be to lower that person's Y score, on average, by 4.00 – 5.00 = -1.0 units relative to the typical score, namely the overall grand mean on Y. The second person has a "background" mean of 2.0. The effect of this individual's background is to lower his or her score, on average by 2.00 – 5.00 or -3.00 units relative to the typical score represented by the grand mean of Y. If I want to, I can remove the effects of these background influences from the data by adding the opposite of the background effect for a person to that person's scores at each time point. For example, to nullify the effects of the first individual's background (which is to lower his or her score by -1 unit), I add one unit to that person's scores at each time point. With this increase, the separate occasion specific scores for the

individual go from 3, 4, 2, 6, 5 to 4, 5, 3, 7, 6 and the column Mean of these new scores for the individual is 5.0. To nullify the effects of the second individual's background (which was -3), I add three units to that person's scores at each time point. The scores transform from 3, 4, 1, 2, 0 to 6, 7, 4, 5, 3 and the column Mean of these new scores for this individual also becomes 5.0. I can repeat this process for every person for both Y and M separately (note: the grand mean of M was 6.0). Table 16.4 presents the background-nullified scores:

Table 16.4: Background Nullified Scores for M and Y

	Scores on Y						Scores on M					
Person	T1	T2	T3	T4	T5	Mean	T1	T2	T3	T4	T5	Mean
1	4	5	3	7	6	5.00	6	7	5	7	5	6.00
2	6	7	4	5	3	5.00	7	8	5	5	5	6.00
3	5	7	3	5	5	5.00	6	8	4	6	6	6.00
.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.
200	7	6	7	4	1	5.00	7	7	7	5	4	6.00

Note that for the Y scores and looking across the entire matrix of Y scores, there is less variability among the scores than for the original scores. This should not be surprising because I have removed a source of variability from the data, namely the across-person background effects. The same is true for the M scores. Note also that the column Means now all equal the grand mean of Y or M respectively. This is again because I have nullified the across-person background effects from the data. The variability in the nullified Y and M matrices reflects only within-person variability across time plus random noise. The between-person variability has been purged.

With this new data that reflects only within-person variability, I can regress the nullified Y scores onto the corresponding nullified M scores. The result will be a within-person regression/path coefficient that estimates the causal link between M and Y over time rather than across individuals. Interestingly, this approach eliminates the confounding effects of all time-invariant confounds like biological sex, ethnicity, and SES because they have been removed from the data. With the nullified data, I can document how Y varies as a function of M as M changes across time.

The above is a simplification of the analytics involved and the ways of segregating between-person variation from within-person variation but it provides an intuitive sense

of the different ways of documenting causal links between variables in panel models, a between-person approach and/or a within-person approach. Which index do you find more appealing as reflective of the causal link between M and Y? Those who favor the within-person coefficient note that it is free of time-invariant confounds and that it analyzes scenarios where changes in M have truly taken place across time. In between-person analyses, they argue, changes in M have not necessarily occurred; instead, we are comparing people on Y who score one unit higher on M with people who score one unit lower on M to make a causal inference. "Where is the change?" the critic asks. Those who favor between-person coefficients acknowledge the problem of time-invariant confounds but they do their best to control for them statistically. The wider coverage of scores across M and Y is seen as a strength as is the focus on individual differences.

I often find in my research that the magnitude of the two types of coefficients are comparable which then increases my confidence in them. But in the final analysis, the coefficients reflect somewhat different phenomena and there will be cases where one is more appropriate depending on one's goals and orientations (Hamaker, 2023). A classic example of opposing dynamics between them is for the effect of physical exercise on the incidence of heart attacks in adults. On a between-person level, there is a negative association between the amount that people exercise and the occurrence of heart attacks; people who exercise more than other people tend to be less likely to have a heart attack than people who exercise less. However, within a given person, heart attacks are more likely to occur when people are exercising, so there is a positive link between the amount of exercise and heart attacks within people. A panel model would find opposite signed causal coefficients in this case. Such dynamics probably are the exception. For example, getting wet due to rain exposure is likely the same when rain exposure comparisons are made across people (it rains on one person but not another person) or within-persons over time (it rains on a person one day but not on another day). In later sections, I describe methods to isolate within-person and between-person coefficients; see Hoover et al. (2019) for a discussion of conditions that promote disparate coefficients.

Parenthetically, a common version of the mathematical model for panel data that captures the above ideas is as follows, expressed here using population notation and where T is the number of assessments made across time:

$$Y_{it} = \mu_t + \beta X_{it} + \gamma Z_i + \alpha_i + \varepsilon_{it} \qquad [16.6]$$

where μ_t is an intercept that can be different at each time point, X_{it} is a vector of predictors that change over time across the measurement occasions, Z_i is a vector of time-invariant predictors, ε_{it} is a random disturbance term, and α_i represents all differences between individuals that are stable over time (time-invariant) and not otherwise

accounted for by Z_i. The β and γ are vectors of regression/path coefficients for the X and Z predictors. There are variants of this model that I consider later.

The Cross-Lagged Panel Model

One of the most well-known panel models for analyzing causal links between two variables is the classic **cross-lagged panel model** (CLPM; Greenberg & Kessler, 1982; Ormel et al., 2002; Kenny, 2005; Muthén & Asparouhov, 2024; Muthén, Asparouhov & Witkiewitz, 2024; Zyphur, et al. 2020a; Zyphur, et al. 2020b). You likely have encountered a version of it in your readings. The model comes in different forms but it typically embraces autoregressive effects, reciprocal causality, lagged effects and contemporaneous effects in one form or another. It also is frequently used in mediation analyses (see Maxwell & Cole, 2007; Maxwell, Cole & Mitchell, 2011). Figure 16.9 presents a classic variant of the model for the two variable case involving a single mediator and a single outcome measured at three time points (posttreatment). There are four first order autoregressive effects, p_1 through p_4. There are no second order autoregressive effects. There are two lagged effects of M on Y, p_5 and p_6, that reflect causal links of M on Y over time. Similarly, there are two lagged effects of Y on M, p_7 and p_8, that reflect causal links of Y on M over time. As such, the model assumes reciprocal causality between M and Y but it is not contemporaneous; there is a lag between them. Similarly, within times 2 and 3, there are correlated disturbances that are included to adjust for the effects of within-time unmeasured confounds between M and Y. Note that there are assumed to be no time-varying unmeasured confounds because of the absence of a correlation between d1 and d2 and between d3 and d4. The relationship between M and Y at time 1 is treated as "unanalyzed," which is why there is only a curved arrow between these variables signifying a correlation between them. This is because the path coefficients for them cannot be estimated because we lack measures of the autoregressive functions leading up to them that need to be statistically controlled to make causal estimates. As such, the "meat" of the model is the various path coefficients, which are estimable with the M_{t1} and Y_{t1} variables treated as exogenous.

Although this model is popular, there are non-trivial problems. The most flagrant one is that there are multiple "equivalent" competing models that produce identical chi square values making the models indistinguishable at this level. Muthén and Asparouhov (2024) identify five such models shown in Figure 16.10. In this figure, the circled disturbances with arrows emanating towards their respective endogenous variables are omitted but they are present and notated using short arrows without the circles. Some of the models have letters associated with the path coefficients which reflect equality constraints imposed to avoid under-identification. If paths share the same letter, this

means they are constrained to be equal in value. Note that some of the models specify contemporaneous reciprocal causality while others specify lagged reciprocal causality and others specify both. None of the models specify unmeasured time-varying confounds, the implication being that they are thought to be trivial in magnitude. The conundrum is that the problem of equivalent models must be addressed for cross-lagged modeling.

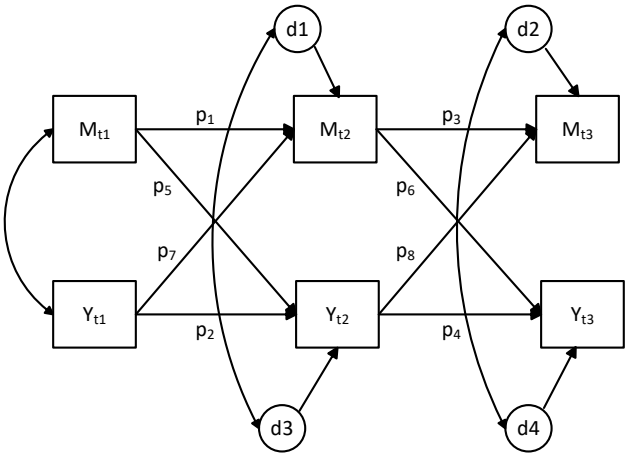


FIGURE 16.9. Classic two variable three wave cross-lagged model

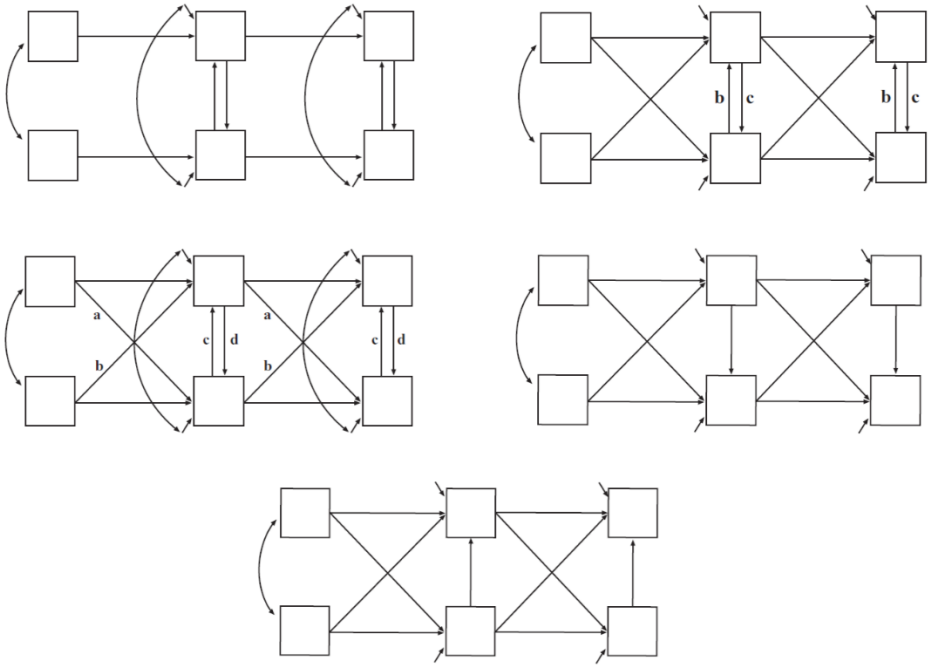


FIGURE 16.10. Competing "equivalent" models for the classic cross-lagged model

For an RET, the classic cross-lagged panel model can be modified to include a treatment variable per Figure 16.11. The intervention is implemented at time 0 so that the mediators and the outcome measures at *t0* reflect baseline measures.

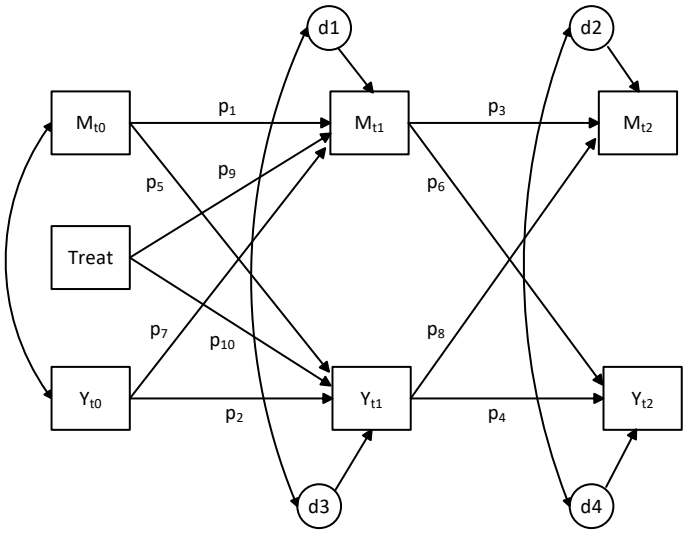


FIGURE 16.11. Two variable three wave cross-lagged model for an RET

The number of cross-lagged model variants that one can potentially invoke seems almost endless, which brings us to another problem with using such models as a blueprint for RET analyses, namely that of model complexity. Model complexity in RETs with many waves can be challenging because more complex models often multiply the number of "equivalent" model competitors to contend with. More waves also increase the number of model variables one must deal with to the point the system can become unstable especially given typical RET sample sizes. Suppose I have three mediators instead of one, each of which is potentially impacted by a treatment condition at baseline. In theory, the treatment condition may have both direct and indirect effects in the model on each of the various mediators across time as well as on the outcome. Suppose instead of three time points, I have five time points in the three mediator model with complex causal relationships among the mediators. Suppose further there are second and third order autoregressive effects for some variables. I undoubtedly also will have a host of covariates for each endogenous variable to control for, especially for the M to Y links.

The situation becomes even more complex if I have multiple indicators of variables in order to address measurement error through the formation of latent variables. Model complexity can quickly get out of hand. Many of the academic treatments of cross-lagged

panel models in statistical textbooks fall short in realistic settings. Program evaluators need to make compromises when testing models in applied settings in order to deal with model complexity. The issue becomes what type of compromises to make.

The blueprint and strategies for compromise that I describe in this chapter are first and foremost guided by the three central questions of an RET elaborated in previous chapters, namely (1) does the intervention meaningfully affect the target outcome, (2) does the intervention meaningfully affect the mediators that supposedly affect the outcome, and (3) do the targeted mediators, in fact, meaningfully impact the outcome. I address these questions in RET designs that have multiple longitudinal assessments, which separates my current focus from prior chapters. My approach deals with model complexity by making liberal use of limited information SEM (LISEM) as opposed to full information SEM (FISEM) and by focusing on generating answers to the above three questions. I introduced LISEM in Chapter 8 and have touched on it in several prior chapters to this one. The models I use in practice generally have three to five mediators in them. There may be scenarios where you have more mediators than this and you might need to prioritize the particular mediators you focus on, a matter I address in Chapter 17. As with previous chapters, I am not so concerned with evaluating omnibus mediation effects, although I always obtain perspectives on them via joint significance tests and evaluating the strengths of the separate links within each mediational chain.

A final issue I want to mention about cross-lagged panel models is the need to recognize the role that time-invariant and time-varying constructs play in them. A striking example is provided by Lucas (2023) who shows how simple correlations between stable, time-invariant variables can masquerade as cross-lagged effects over time. Figure 16.12a is a causal diagram that has two time-invariant, stable latent variables, X and Y, that are correlated 0.70 but are not causally related. The observed variables X1 and X2 are measured at two points in time as are the variables Y1 and Y2. There, of course, are disturbance terms for X1, X2, Y1, and Y2 that reflect factors other than the stable latent X and Y that impact the X1, X2, Y1, and Y2 variables. The disturbances reflect occasion-specific causes that randomly shift scores upward or downward on the observed variables. In a simulation, Lucas set the variance of the latent X and latent Y to both equal 1.0 and the path coefficients linking them to the time-specific assessments equal to 0.70. The disturbance variances were set to 0.50 and the variances of X1, X2, Y1, and Y2 all were set to 1.0. Lucas treated this model as the underlying data generating process and generated data that fully conformed to it. However, he then analyzed the generated data using a blatantly misspecified two-variable, two-wave cross-lagged panel model. The results are in Figure 16.12b.

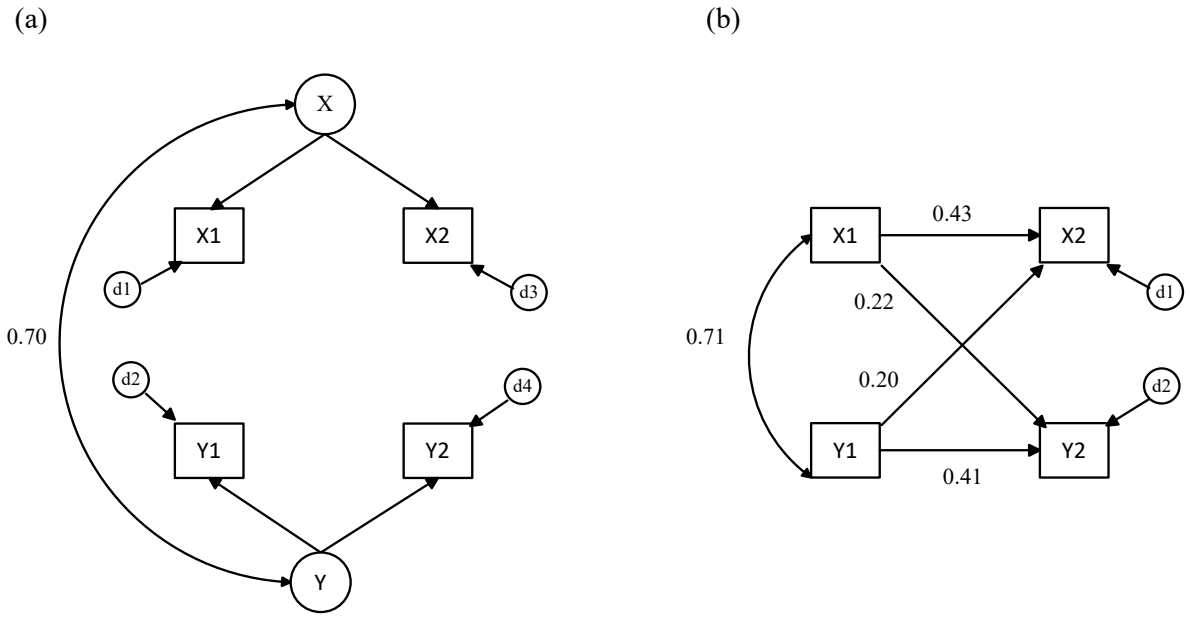


FIGURE 16.12. Equivalent models

In terms of model fit, the misspecified model yielded perfect fit to the data. Note that there is support in Figure 16.12b for reciprocal causal links between X and Y at the observed level despite the fact there are no true across-time causal effects at play in the data generating process. In the final analysis, the traditional cross-lagged panel model cannot distinguish between associations that occur at the stable, time-invariant level (Figure 16.12a) from those that occur at the time-varying level. The traditional analytic approach needs to be modified to take time-invariant process into account. Ironically, if I treat the model in Figure 16.12b as the data-generating process rather than the model in Figure 16.12a, the model in Figure 16.12a would produce a perfect fit to the data generated by it. Lucas (2023) argues that the analysis of panel data needs to take into account both time-invariant and time-varying phenomena which translates into dealing in one way or another with between-person and within-person variability.

In sum, the CLPM has been popular in the social sciences and often is the "model of choice" for approaching panel data in statistical textbooks. This is despite the fact that the approach has been characterized in less than flattering terms in the statistical literature, such as in the article by Lucas (2023) titled "*Why the cross-lagged panel model is almost never the right choice.*" For RETs, I believe there usually are more appropriate model forms and analytic methods that can be invoked to map the impact of mediators on outcomes although the cross-lagged panel model always is a possible candidate.

Concluding Comments on Core Concepts in Longitudinal RETs

This chapter section has established multiple foundations for longitudinal modeling. Most important is appreciating the fact that causal lags are multi-faceted in nature. Causal inferences will almost always be constrained by the particular time intervals used in a study as well as other study facets. Just as we need to be cautious about generalizing inferences to other populations, settings, and contexts, so too must we be cautious about generalizing inferences to other time lags. If you desire to generalize your results to other time lags, then your research design should reflect strategies that permit such generalizations. Gollob and Reichart's (1991) concepts of time-specific direct and indirect effects and time-specific omnibus mediation effects are fundamental.

Longitudinal models often address reciprocal causality between mediators or between mediators and outcomes, sometimes contemporaneously and sometimes longitudinally with notable lags between assessments. Knowing how to model contemporaneous reciprocal causality is important because longitudinal RETs sometimes include phenomena with short causal lags that functionally operate contemporaneously; or the RETs might focus on time periods in which earlier reciprocal dynamics have played themselves out and hence are meaningfully captured by loop-based modeling with concurrent measures.

Dealing with confounds is critical in all RETs and you need to make choices about which confounds to measure and which ones to leave unmeasured (see Chapter 2). There are time-invariant confounds and time-varying confounds to think about as well as covariates that when measured and controlled can increase your statistical power and precision. Instrumental variables are a useful tool to have at hand for dealing with confounds and thinking through their role when planning your study is important.

Autoregressive and lagged effects inevitably come into play in many longitudinal studies and you need to carefully consider their role in your RET. This includes thinking about the resultant statistical controls that may be necessary to allow effective autoregressive modeling. Finally, your primary focus can be on within-person coefficients, between-person coefficients, or both. What are your priorities relative to these types of coefficients? This needs to be thought about a priori.

In many longitudinal studies you will be faced with non-trivial model complexity. You need a blueprint for dealing with it. SEMers often seek to analyze a single, grand model that simultaneously addresses all core questions of interest and then evaluate the viability of the model and its coefficients using full information estimation SEM. Although there are advantages to doing so, there also are disadvantages (see Chapter 8 for a discussion of these issues). Limited information SEM can be an attractive alternative. To be sure, LISEM is still guided by a single, grand SEM model. However,

the analytics are strategically subdivided into different components with the LISEM results then pieced together to yield perspectives on the overall model.

At this point it is worth re-iterating a point I made in Chapter 7. In SEM, we test casual models not prove them. Unambiguously proving causality for the types of constructs we study in the social and health sciences is almost impossible. In SEM, we (1) posit a causal model that we think may be operating in a given context, (2) we specify the predictions the model makes about how the data that we collect to evaluate the model should pattern themselves, (3) we analyze the data to determine if they do indeed pattern themselves in the predicted fashion, (4) we reject the model if the data do not pattern themselves in the predicted ways. If the data pattern themselves as predicted by the model, our confidence in the viability of the model increases, but this increase in confidence is dependent on the quality of the research design we used. If the model proves to be consistent with the data, then we interpret the parameter estimates of the model to gain further insights into the phenomena we are studying. This general logic applies to observational data, experimental data, and indeed most any quantitative study one conducts, be it cross-sectional or longitudinal. We evaluate the viability of causal models; we do not prove causality.

In the next few sections, I further explicate statistical foundations for analyzing longitudinal RETs with multiple follow-ups. I first consider a method in the panel design literature known as **fixed effects modeling** and then comment on a method known as **random effects modeling**. After introducing these approaches, I describe the methods of GLM based within-between modeling and generalized estimating equations (GEE). I then propose a tentative RET framework for data analysis using a detailed numerical example that applies the framework. Future sections of this chapter consider a wide range of additional SEM based methods for analyzing longitudinal RETs.

SEM-BASED FIXED EFFECTS PANEL MODELS

In RETs, I often use SEM-based fixed effects panel methods described by Allsion et al. (2017) to examine the relationship between mediators and outcomes in longitudinal models. The method primarily focuses on estimating within-person path coefficients for M→Y links but it can be expanded to analyze certain types of between-person effects as well. I ultimately come at matters from both between- and within-person perspectives but my focus here is on the fundamentals of SEM-based fixed effects panel models.

Consider a study in which a mediator, M, and an outcome, Y, are measured at five points in time after an intervention. In SEM-based fixed effects modeling, the focus for estimating a causal coefficient for Y autoregression is how variation in Y at

time *t* predicts variation in Y at time *t+1* across the 5 occasions. This is accomplished by analyzing the measures of Y*t* and Y*t+1 within individuals*. The same is true when examining a causal coefficient linking M to Y. The focus is on estimating the regression coefficient for Y on M *within individuals* and across occasions. If I tell you I applied SEM-based fixed effects modeling and the estimated within-individual coefficient for the regression of Y on M was 0.50, this means that for a given individual, if M changes 1 unit from one occasion to the next occasion, there likely will be a concomitant change of 0.50 in Y across those occasions.

The SEM-based fixed effects approach offers a different perspective on estimating the causal impact of Y*t* on Y*t+1* or of M on Y than traditional cross-sectional approaches because of its within-individual, across-occasion focus. In doing so, *all* stable, time invariant confounds become irrelevant. Factors like biological sex, parental upbringing, ethnicity and the like cannot be a confound in such analyses because they are held constant when working exclusively with within-individual variation. Many methodologists prefer the fixed effects statistical approach to characterizing causality.

I am sometimes asked by students about when they should *a priori* include autoregressive causal effects in their causal models. My response is that they should consider the possibility of doing so but with the mindset that if they control for all possible across-individual, time-invariant stable confounds, do they still believe there will be an association or influence of Y*t* on Y*t+1*. If the answer is yes, then perhaps include the autoregressive path in the model. If the answer is no, then unless there are strong methodological reasons for doing so, the effect probably should be omitted.

The same mindset should be brought to bear when estimating causal coefficients between M and Y. Factors that are relatively stable over the study period, like SES, place of residence, personality traits, genetic predispositions, and cognitive abilities, all represent confounds that can contaminate estimates of the M→Y link by being common causes of both M and Y. The power of SEM-based fixed effects modeling is that all such confounds, whether measured or unmeasured, are controlled. My presentation of this method of analysis begins with simple and somewhat restrictive models using continuous outcomes but eventually I extend matters to more complex designs.

The Basics

The first model I present is the classic fixed effects regression model that evolved outside of SEM contexts. Allison's approach evolves from this tradition. The fixed effects regression models has its roots in econometrics and seeks to document within-person coefficients in longitudinal panel models. One simple but cumbersome approach to estimating M→Y within-person path coefficients for continuous outcomes requires that

we first format our data in long format rather than wide format. Let the number 1 refer to the immediate posttest, 2 refer to a 3 month follow-up, 3 to a 6 month follow-up, and 4 to a 9 month follow-up. Let the variable Y be the outcome, MA a mediator and MB a second mediator. ID is a case ID variable and SEX is a time-invariant measure of biological sex at birth. Wide format data, which is traditional, are structured as follows (I show only the first three cases):

ID	Y1	Y2	Y3	Y4	MA1	MA2	MA3	MA4	MB1	MB2	MB3	MB4	SEX
1	10.1	14.2	14.4	14.3	8.1	10.2	10.3	10.1	5.2	7.5	7.9	7.1	0
2	10.1	14.2	14.3	14.4	8.5	10.6	10.7	10.8	5.9	7.0	7.1	7.2	1
3	10.8	14.9	14.0	14.1	8.2	10.3	10.4	10.5	5.6	7.7	7.8	7.9	0

The same data in long format appears as follows:

ID	TIME	Y	MA	MB	SEX
1	1	10.1	8.1	5.2	0
1	2	14.2	10.2	7.5	0
1	3	14.4	10.3	7.9	0
1	4	14.3	10.1	7.1	0
2	1	10.1	8.5	5.9	1
2	2	14.2	10.6	7.0	1
2	3	14.3	10.7	7.1	1
2	4	14.4	10.8	7.2	1
3	1	10.8	8.2	5.6	0
3	2	14.9	10.3	7.7	0
3	3	14.0	10.4	7.8	0
3	4	14.1	10.5	7.9	0

In long format, there are 4 rows per participant, one for each time point. A new variable, TIME, is created to reflect the four time points. The time-varying variables (Y, MA, and MB) have unique scores for each time point. The time-invariant variables (SEX) repeat the same score at each time point for a given participant. If I have 200 study participants at four time points each, the data matrix will have a total of 800 rows. My website has a program called *Wide to long data format* that converts wide format to long format data.

The traditional fixed effects regression approach works with long format data. It adds to the above data a set of dummy variables that distinguish the different individuals. If there are 200 individuals, we create 199 (or *k*-1) such dummy variables. The first dummy variable assigns a 1 to each row in the data matrix that has a score from participant 1 and a 0 to all other rows. The second dummy variable assigns a 1 to each row that has a score from participant 2 and a 0 to all other rows. The third dummy

variable assigns a 1 to each row that has a score from participant 3 and a 0 to all other rows. And so on. I then regress the 800 Y scores onto the 800 MA and 800 MB scores and the 199 dummy variables. The inclusion of the dummy variables in the prediction equation has the effect of removing all between-person variability so that all that remains is within-person variability plus error. The coefficients of interest in this analysis are the ones for MA predicting Y and MB predicting Y; they are within-person causal estimates.

The above strategy is quite inefficient. SEM offers a more elegant and flexible approach to fixed effects panel modeling. Figure 16.13 presents an influence diagram for a single mediator and a single outcome with T = 3 to illustrate the SEM-based logic.

This model assumes the effect of M on Y is unidirectional and contemporaneous. I relax this assumption later. Each of the M→Y path coefficients have a common label (p_1) meaning they will be constrained to be equal during the analysis. The estimated coefficient for p_1 in the output tells us for every one unit that M changes over time, how many units Y is expected to change within persons. The M at the three time points are allowed to be correlated. There may be autoregressive causal effects between the M but for now, I do not model them. I instead treat the correlations among the M as causally "unanalyzed." Doing so does not bias the estimates of p_1, which I am most interested in.

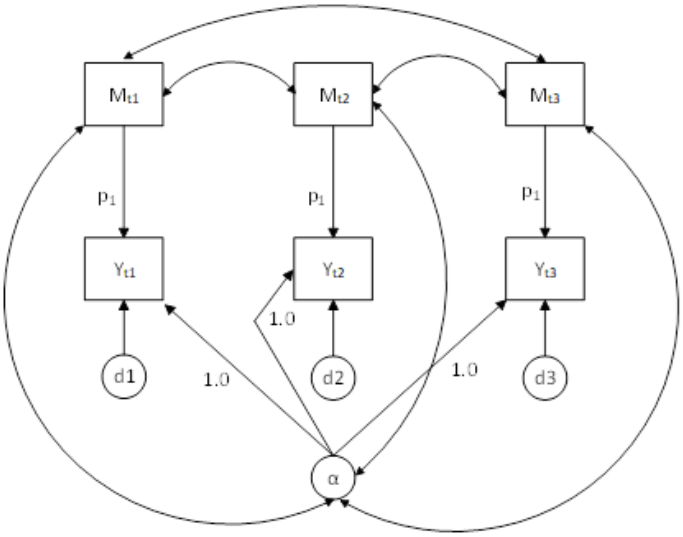


FIGURE 16.13. Traditional fixed effect regression expressed using SEM.

The three disturbance terms for Y are assumed to be uncorrelated across time, which translates into the presumption that there are no meaningful time-varying confounds. I show you later how to modify the model to address such confounds.

Although not shown in the diagram, I also will constrain the three disturbance variances to be equal. Later, I will relax this assumption as well. Finally, the model includes a latent variable labeled α with causal paths emanating to each of the Y variables but with each path fixed to a value of 1.0. This is a somewhat counter-intuitive programming strategy that accomplishes what the individual-based dummy variables in the prior approach accomplished, namely it controls for all time-invariant across-individual confounds, measured or unmeasured. For explication of why this works, see Andersen (2022). Note also that the α latent variable is allowed to correlate with the various M variables in the model, which is an important property, as explained later.

To make the above concrete, I generated hypothetical data from an RET that had five post intervention assessments, the immediate posttest and four additional follow-ups at 3 month intervals, namely 3, 6, 9 and 12 months. No baseline measures were obtained, i.e., it was a two-group posttest only design. In addition to the outcome Y, two mediators were assessed, MA and MB at each time point as well as a time varying covariate, COV. All of the measures had metrics from approximately 0 to 10 with standard deviations ranging from 2 to 3. My goal is to use SEM-based fixed effects panel modeling to obtain an estimate of the within-person causal coefficient linking each mediator to the outcome controlling for COV. In the numerical examples I consider later in the chapter, I will conduct a full-fledged analysis of the three key questions of an RET. One of the tools I invoke, among others, is SEM-based fixed effects panel modeling. The current treatment introduces SEM-based fixed effects panel modeling in a more narrow way.

Parenthetically, including time-varying covariates like COV in the model is important for controlling for time-varying confounds which then reduces the need to introduce time-varying correlated disturbances into the model such as a correlation between d1 and d2 in Figure 16.13. To be sure, such correlated disturbances in traditional modeling also can be caused by time-invariant common causes, but the SEM-based fixed effects panel model effectively controls for time-invariant confounds both measured and unmeasured alike. However, there may be residual disturbance dependencies at work due to time-varying confounds over and above the core predictors in the model. These are controlled by including covariates that represent them in the model.

The relevant Mplus syntax is in Table 16.5.

Table 16.5: Mplus Syntax for SEM-based Fixed Effects Panel Models

```
1.  TITLE: SEM-based f1xed effects model I
2.  DATA: FILE = FEexample1.txt;
3.  VARIABLE:
4.    NAMES ARE id za zb ma1 ma2 ma3 ma4 ma5 mb1 mb2 mb3 mb4 mb5
5.    cov1 cov2 cov3 cov4 cov5 y1 y2 y3 y4 y5;
```

```
6.  USEVARIABLES ARE ma1 ma2 ma3 ma4 ma5 mb1 mb2 mb3 mb4 mb5
7.    cov1 cov2 cov3 cov4 cov5 y1 y2 y3 y4 y5 ;
8.  ANALYSIS: ESTIMATOR=MLR ;
9.  MODEL:
10.   alpha BY y1@1 y2@1 y3@1 y4@1 y5@1; !define latent time-invariant var
11.  y1 ON ma1 mb1 cov1 (p1 p2 b1) ; !regress t1 y onto t1 preds
12.  y2 ON ma2 mb2 cov2 (p1 p2 b1) ; !regress t2 y onto t2 preds
13.  y3 ON ma3 mb3 cov3 (p1 p2 b1) ; !regress t3 y onto t3 preds
14.  y4 ON ma4 mb4 cov4 (p1 p2 b1) ; !regress t4 y onto t4 preds
15.  y5 ON ma5 mb5 cov5 (p1 p2 b1) ; !regress t5 y onto t5 preds
16.  y1 y2 y3 y4 y5 (evar);          !estimate disturbance variances
17.  alpha WITH ma1-ma5 mb1-mb5 cov1-cov5 ; !correlate latent var with all Xs
18.  OUTPUT: SAMP STDYX MOD(4) CINTERVAL RESIDUAL TECH4  ;
```

On line 4, I make sure to use variable names that end with sequential numbers corresponding to the time points so that I can take advantage of a programming shortcut I show you shortly. On Line 6, I make sure each variable is ordered sequentially by their end number, again to take advantage of the programming shortcut. Although the variables already were in proper order on the `NAMES ARE` lines and I did not need to reorder anything on the `USEVARIABLES` line, I could do so if necessary.

Line 10 defines the latent α variable that reflects the influence of unmeasured time-invariant determinants of the outcome. I use the `BY` command as a programming shortcut to fix each path from the latent variable to the observed Y to equal 1.0 using the Mplus `@1` convention. Lines 11-15 regress the Y variables at each time point onto the relevant predictors at each time point and assign labels to the three coefficients in each equation, namely `p1`, `p2`, `b1`, using the Mplus convention discussed in prior chapters. Note that the corresponding coefficients in each equation have identical labels which has the effect of imposing an equality constraint on coefficients that share the same label. Line 16 tells Mplus to estimate the disturbance variances and the label `evar` at the end of the line applies to each disturbance term, thereby forcing an equality constraint on these variances by means of having a common label. Line 17 correlates the latent variable with all of the observed predictors using the Mplus programming shortcut I mentioned; the notation MA1-MA5 actually refers to MA1, MA2, MA3, MA4 and MA5, i.e., all variables between MA1 and MA5 inclusive. The variables listed on the left side of `WITH` are correlated with all the variables listed on the right side of `WITH`.

Missing data are handled using the Mplus default of FIML although this example had no missing data. I used a robust maximum likelihood estimator (see Line 8).

The global fit indices of the model all suggested satisfactory model fit. The chi square was 71.04 with df=70, $p < 0.45$; the RMSEA was 0.004 with a 90% confidence interval of 0.00 to 0.019; the p value for close fit was < 1.00; the CFI was 0.99 and the standardized RMR was 0.014. For localized fit, there were no statistically significant

differences between the predicted and observed covariances on a cell-by-cell basis and there were no meaningful modification indices above 4.0. The sample size was N = 1000.

Here are the results for the key unstandardized regression coefficients:

		Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
Y1	ON				
MA1		0.310	0.015	20.118	0.000
MB1		0.304	0.016	19.512	0.000
COV1		0.296	0.016	18.619	0.000
Y2	ON				
MA2		0.310	0.015	20.118	0.000
MB2		0.304	0.016	19.512	0.000
COV2		0.296	0.016	18.619	0.000
Y3	ON				
MA3		0.310	0.015	20.118	0.000
MB3		0.304	0.016	19.512	0.000
COV3		0.296	0.016	18.619	0.000
Y4	ON				
MA4		0.310	0.015	20.118	0.000
MB4		0.304	0.016	19.512	0.000
COV4		0.296	0.016	18.619	0.000
Y5	ON				
MA5		0.310	0.015	20.118	0.000
MB5		0.304	0.016	19.512	0.000
COV5		0.296	0.016	18.619	0.000

Note that the results are identical for each time point. This reflects the equality constraints that I imposed in the program. As I discuss later, we usually do not interpret the coefficients at each time point in their own right. Rather, we focus on the single common coefficient value that characterizes each time point by virtue of the built-in equality constraints. The original fixed effects regression approach that focuses on within-person, across-occasion coefficients was developed by economists who do not use SEM. The mathematics of the regression-based approach they evolved are rather complex and carry with them a host of stringent assumptions. Allison et al. (2016) evolved a strategy that generalizes the econometric approach to SEM and in doing so, brought considerably more flexibility to it. The resulting SEM model that is fit to data is somewhat non-intuitive but ultimately it replicates the well-established approach by economists but with greater flexibility. The mechanics of the SEM approach require that

we introduce equality constraints for the coefficients at the five time points but we only interpret the one value that results as the within-person across-occasions coefficient.

For the MA mediator, the estimated coefficient was 0.310 ± 0.03 (Critical Ratio (CR) = 20.12, $p < 0.05$). Within individuals and across time, for every one unit that MA increases, the outcome Y is predicted to increase 0.310 units holding constant MA and COV. For the MB mediator, the estimated coefficient was 0.304 ± 0.03 (Critical Ratio (CR) = 19.51, $p < 0.05$). Within individuals and across time, for every one unit that MB increases, the outcome Y is predicted to increase 0.304 units.

Traditionally, we do not examine standardized regression coefficients when the focus is on within-person analyses. This is because such coefficients can be misleading given the focus is on within-person variation. Variables are standardized using across-individual standard deviations but with SEM-based fixed effects panel models, it is within-person variability that matters, not between-individual variability.

A portion of the output I often examine is the estimated correlations between the latent α variable and the time-varying predictors. Here are the results for MA and MB:

```
STDYX Standardization
```

		Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
ALPHA	WITH				
MA1		-0.036	0.038	-0.950	0.342
MA2		0.046	0.038	1.231	0.218
MA3		0.039	0.036	1.085	0.278
MA4		-0.014	0.035	-0.403	0.687
MA5		0.075	0.037	2.002	0.045
MB1		0.003	0.036	0.084	0.933
MB2		-0.029	0.035	-0.842	0.400
MB3		0.051	0.038	1.354	0.176
MB4		0.024	0.036	0.666	0.505
MB5		-0.028	0.037	-0.742	0.458

Note that all of the correlations are near zero and statistically non-significant. A feature of the SEM-based fixed effects panel model is that it allows for these correlations to be present. This is not the case for several other models I discuss later, such as the SEM-based random effects model. In the current case, the above meager correlations suggest there likely are no meaningful time-invariant unmeasured confounds between the mediators and the outcome; if I pursue a modeling strategy that does not correct for unmeasured time invariant confounds, at least for the MA→Y and MB→Y relationships, I likely will not be misled. In practice, it is more common for some of these correlations to be meaningfully non-zero and statistically significant which then places an onus on

investigators to measure and statistically control for time-invariant confounds when applying models that do not adjust for the confounds. The presence of such confounds, by contrast, poses no problem for the SEM-based fixed effects panel model.

Relaxing Model Constraints

The SEM-based fixed effects panel model imposed multiple constraints on model parameters. Not all of them are necessary but they do promote model parsimony and were necessary in the early incarnations of fixed effects modeling with traditional non-SEM software. In this section, I show how to relax the constraints and how to test the reasonableness of such relaxations.

Disturbance Variance Constraints

The tested SEM-based model assumed equal disturbance variances for the Y across time. This assumption is not required and we can remove it accordingly. Imposing it yields a more parsimonious model but if the disturbance variances are markedly different, one can always relax the constraint. I re-ran the model without the restriction by removing the `evar` label on Line 16 of Table 16.5. Here is the chi square index for this model:

```
Chi-Square Test of Model Fit

          Value                             67.922*
          Degrees of Freedom                    66
          P-Value                           0.4115
          Scaling Correction Factor         0.9881
            for MLR
```

and here is the result from the original model with the constrained variances:

```
Chi-Square Test of Model Fit

    Value                                   71.038*
          Degrees of Freedom                    70
          P-Value                           0.4429
          Scaling Correction Factor         0.9870
            for MLR
```

I used the program on my website titled *Scaled chi sqr difference test* to conduct a formal nested chi square difference test between the two models. The scale corrected chi square difference was 3.10 with 4 degrees of freedom and yielded a p value < 0.55. I therefore cannot confidently reject the null hypothesis of equal population disturbance variances, suggesting my original model with the constraint probably is reasonable.

The above is an omnibus test of the equal disturbance variance constraint. I can obtain more detailed contrasts using the MODEL CONSTRAINT command in Mplus. Table 16.6 presents the relevant syntax.

Table 16.6: Mplus Syntax Using the Model Constraint Command

```
1.  TITLE: SEM-based f1xed effects model I
2.  DATA: FILE = FEexample1.txt;
3.  VARIABLE:
4.    NAMES ARE id za zb ma1 ma2 ma3 ma4 ma5 mb1 mb2 mb3 mb4 mb5
5.    cov1 cov2 cov3 cov4 cov5 y1 y2 y3 y4 y5;
6.  USEVARIABLES ARE ma1 ma2 ma3 ma4 ma5 mb1 mb2 mb3 mb4 mb5
7.    cov1 cov2 cov3 cov4 cov5 y1 y2 y3 y4 y5 ;
8.  ANALYSIS: ESTIMATOR=MLR ;
9.  MODEL:
10.   alpha BY y1@1 y2@1 y3@1 y4@1 y5@1; !define latent time-invariant var
11.  y1 ON ma1 mb1 cov1 (p1 p2 b1) ; !regress t1 y onto t1 preds
12.  y2 ON ma2 mb2 cov2 (p1 p2 b1) ; !regress t2 y onto t2 preds
13.  y3 ON ma3 mb3 cov3 (p1 p2 b1) ; !regress t3 y onto t3 preds
14.  y4 ON ma4 mb4 cov4 (p1 p2 b1) ; !regress t4 y onto t4 preds
15.  y5 ON ma5 mb5 cov5 (p1 p2 b1) ; !regress t5 y onto t5 preds
16.  y1 (evar1) ; y2 (evar2) ; y3 (evar3) ;  y4 (evar4) ;  y5 (evar5) ;
17.  alpha WITH ma1-ma5 mb1-mb5 cov1-cov5 ; !correlate latent var with all Xs
17a. MODEL CONSTRAINT:
17b. NEW (diff1-diff10) ;
17c. diff1 = evar1-evar2;
17d. diff2 = evar1-evar3;
17e. diff3 = evar1-evar4;
17f. diff4 = evar1-evar5;
17g. diff5 = evar2-evar3;
17h. diff6 = evar2-evar4;
17i. diff7 = evar2-evar5;
17j. diff8 = evar3-evar4;
17k. diff9 = evar3-evar5;
17l. diff10 = evar4-evar5;
18.  OUTPUT: SAMP STDYX CINTERVAL RESIDUAL TECH4  ;
```

I replaced Line 16 with a command that gave distinct labels to the five disturbance variances in order to remove the equality constraint. I removed the request for modification indices from the OUTPUT line because Mplus does not allow them with use of the MODEL CONSTRAINTS command. In Line 17b I specify there will be 10 new contrasts by expressing labels for them using Mplus hyphenation shorthand. Lines17c through 17l define the contrasts. From the Mplus output, here are the estimated unstandardized disturbance variances for each Y:

```
                                                    Two-Tailed
                    Estimate       S.E.  Est./S.E.    P-Value
Residual Variances
    Y1                 3.733      0.209     17.878      0.000
    Y2                 3.964      0.205     19.326      0.000
    Y3                 4.167      0.207     20.109      0.000
    Y4                 3.894      0.206     18.895      0.000
    Y5                 4.170      0.207     20.126      0.000
```

The disturbances are not all equal but they are very close in value to one another; their differences could just reflect sampling error. Here are the 10 statistical contrasts:

```
                                                    Two-Tailed
                    Estimate       S.E.  Est./S.E.    P-Value

New/Additional Parameters
    DIFF1             -0.230      0.298     -0.772      0.440
    DIFF2             -0.434      0.297     -1.460      0.144
    DIFF3             -0.160      0.288     -0.556      0.578
    DIFF4             -0.436      0.296     -1.472      0.141
    DIFF5             -0.204      0.305     -0.667      0.504
    DIFF6              0.070      0.288      0.242      0.808
    DIFF7             -0.206      0.293     -0.703      0.482
    DIFF8              0.274      0.295      0.927      0.354
    DIFF9             -0.002      0.294     -0.007      0.994
    DIFF10            -0.276      0.294     -0.937      0.349
```

Looking at the last column, none of the contrasts were statistically significant ($p < 0.05$). All of the differences in the `ESTIMATE` column appear to be small in magnitude.

Focused tests of variance differences can be problematic so it probably is best to replicate the above using bootstrapping. This involves changing the estimator from `MLR` to `ML` on Line 8 of Table 16.6 and then adding `BOOTSTRAP=2000` on the `ANALYSIS` command, like this:

```
ANALYSIS: ESTIMATOR=ML ; BOOTSTRAP=2000 ;
```

and also changing `CINTERVAL` on the output line to `CINTERVAL(BOOTSTRAP)`. The results for the bootstrap analysis were comparable to the original analysis.

Path Coefficient Constraints

The SEM model also used a constraint of equal M→Y path coefficients at each time point. This usually is a reasonable assumption because the contemporaneous effect of a mediator on an outcome often will not be impacted by the particular points in time the variables are assessed. But such is not always the case. An advantage of imposing the

constraint is that you can summarize the effect of M on Y across time using a single coefficient. However, if that effect differs at some points in time versus others, you will want your model to accommodate this fact given that it makes substantive sense. Of course, you do not have to remove the coefficient equality constraint for every time point. You can remove it for different subsets of time points, as appropriate. Theory should be your guide. In this case, the within-person changes depending on the time block in which invariance is found; changes in M predict changes in Y by the value of the path coefficient within the specific time window to which the equality constraints apply. If the coefficient is unique to a single time point, then it applies to the more narrow (unknown) time window surrounding that point.

To remove the coefficient equality constraints in the model as programmed in Table 16.5, I remove the common labels between the coefficients across time but still assign labels to them for reasons I provide shortly; however, the labels are all unique. Here are the changed Lines from Table 16.5:

```
y1 ON ma1 mb1 cov1 (p1a p2a b1a) ;
y2 ON ma2 mb2 cov2 (p1b p2b b1b) ;
y3 ON ma3 mb3 cov3 (p1c p2c b1c) ;
y4 ON ma4 mb4 cov4 (p1d p2d b1d) ;
y5 ON ma5 mb5 cov5 (p1e p2e b1e) ;
```

When I fit this model, here is the global chi square statistic that resulted:

```
Chi-Square Test of Model Fit

          Value                             54.891*
          Degrees of Freedom                   58
          P-Value                           0.5916
          Scaling Correction Factor         0.9886
```

The chi square from the original model was

```
Chi-Square Test of Model Fit

    Value                             71.038*
          Degrees of Freedom                   70
          P-Value                           0.4429
          Scaling Correction Factor         0.9870
```

I used the program on my website *Scaled chi sqr difference test* to conduct a formal nested chi square difference test between the two models. The scale corrected chi square difference was 16.18 with 12 degrees of freedom and yielded a p value < 0.19. I therefore

cannot confidently reject the null hypothesis of equal M→Y population coefficients, suggesting my original model with the equality constraint probably is reasonable.

As with the disturbance variances, the above is an omnibus test of the equal coefficient constraint. I can obtain more detailed contrasts using the same `MODEL CONSTRAINT` strategy I outlined for the disturbance variances (I do not show the syntax here but it is a straightforward generalization of what I did with the disturbance variances). None of the pairwise contrasts in this analysis were statistically significant. For completeness, here are the path coefficients that resulted in the unconstrained model:

		Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
Y1	ON				
MA1		0.375	0.035	10.766	0.000
MB1		0.292	0.035	8.366	0.000
COV1		0.279	0.035	8.023	0.000
Y2	ON				
MA2		0.265	0.034	7.713	0.000
MB2		0.317	0.034	9.354	0.000
COV2		0.295	0.034	8.700	0.000
Y3	ON				
MA3		0.319	0.035	9.001	0.000
MB3		0.308	0.035	8.762	0.000
COV3		0.356	0.034	10.589	0.000
Y4	ON				
MA4		0.346	0.034	10.072	0.000
MB4		0.318	0.037	8.628	0.000
COV4		0.241	0.036	6.611	0.000
Y5	ON				
MA5		0.250	0.032	7.818	0.000
MB5		0.283	0.035	8.103	0.000
COV5		0.306	0.035	8.642	0.000

You can see that for any given predictor (e.g., MA), the coefficients had quite similar coefficients at each of the five time points.

Time Invariant Variable Constraints

A third set of constraints I made in the original model was to fix the path coefficients emanating from the latent variable α to all equal 1.0. These coefficients reflect the effect of the sum total of all the unmeasured time invariants variables on Y at each time point.

The model constraint implies that these effects are the same for Y at each point in time. But perhaps they are not. Perhaps the unmeasured time invariant variables have a larger impact on Y at some time points than others. If I have a reasonable theory that leads me to such a conclusion, then I can relax the equal influence constraint to evaluate it. To do so, I modify Line 10 in Table 16.5 as follows:

```
alpha BY y1@1 y2 y3 y4 y5;
```

For the latent variable to be identified, I need to give it a metric and I satisfy the need by identifying one of the indicators as the reference indicator, in this case y1. I then remove the `@1` notation for the remaining indicators to allow these coefficients to be freely estimated. The alpha latent variable is scaled (roughly) in the metric of y1. When I fit this model, here is the global chi square statistic that resulted:

```
Chi-Square Test of Model Fit

          Value                             69.128*
          Degrees of Freedom                    66
          P-Value                           0.3723
          Scaling Correction Factor         0.9871
            for MLR
```

The chi square from the original model was

```
Chi-Square Test of Model Fit

    Value                             71.038*
          Degrees of Freedom                70
          P-Value                       0.4429
          Scaling Correction Factor     0.9870
            for MLR
```

I again used the program on my website titled *Scaled chi sqr difference test* to conduct a formal nested chi square difference test between the two models. The scale corrected chi square difference was 1.91 with 4 degrees of freedom and yielded a p value < 0.76. I therefore cannot confidently reject the null hypothesis of equal α coefficients across time for predicting Y, suggesting my original model with the constraint probably is reasonable.

Here are the coefficients that resulted from the unconstrained model:

		Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
ALPHA	BY				
Y1		1.000	0.000	999.000	999.000
Y2		1.021	0.072	14.113	0.000
Y3		1.081	0.078	13.920	0.000
Y4		1.003	0.067	14.873	0.000
Y5		1.060	0.079	13.468	0.000

All of the coefficients were quite similar. The lower the value of the estimate, the less the impact of the unmeasured time invariant variables considered as a totality on Y at a given time point. As with the disturbance variances and path coefficients, I can obtain more detailed contrasts using the `MODEL CONSTRAINT` strategy.[6]

In sum, fixed effects regression models evolved from a modeling framework that imposed numerous constraints on the modeling process. The translation of the approach into SEM gives it considerable flexibility. In the absence of strong theory to guide constraint relaxation, one often starts with a parsimonious model that imposes the constraints I identified. If the model fits poorly or if modification indices suggest one or more of the constraints be relaxed, then one can do so. The most complex model is one that relaxes all of the constraints. It is statistically viable, just complicated to process.

Concluding Comments on SEM-based Fixed Effects Panel Models

SEM-based fixed effects panel modeling is a powerful tool for potentially analyzing mediator-outcome relationships in RETs with many time points. The above introduction to the method only scratches the surface of its uses. SEM-based fixed effect panel models also can be used with latent variables that have multiple indicators to adjust for measurement error, they can be used with clustered data, with binary and count outcomes and they can be amended to include measured time-invariant variables as well. I include a document on my webpage titled *Additional RET Applications of SEM-based Fixed Effects Modeling with Panel Data* that discuss many of these extensions as well as other topics related to SEM-based fixed effects modeling in RETs.

Although it has many strengths, one also must keep in mind the limitations of the method. One shortcoming is that the latent α variable represents an entire package of unspecified time-invariant variables and we do not really know the contents of that package. We do not know which parts of the time-invariant package affect the outcome nor do we know the underlying dynamics that are operating between the variables in the

[6] For some possible restrictions in relaxing these constraints, see Murayama and Gfrörer (2024).

package. Nevertheless, we *do* know that modeling α using SEM-based fixed effects panel modeling often improves our estimates of the other parts of the model even though we are forced to make statements only about time-invariant collectives. Interestingly, there are strategies you can use to bring specific time-invariant variables into SEM-based fixed effects panel modeling. I discuss these strategies in the document *Additional RET Applications of SEM-based Fixed Effects Modeling with Panel Data* on my website.

Fixed-effects estimates derive from changes in variables over time and can be of limited value if individuals do not change across time on those variables. As well, fixed-effects panel estimates can be less reliable when the number of time periods is small. In general, when using SEM-based fixed effects panel modeling you will want to have a minimum of three waves of data but usually more. If you have a very large number of time points via the use of an intensive longitudinal design, such as the collection of diary data or daily recorded physical reactions, the data likely will be better analyzed using the dynamic structural equation modeling methods described below. Most applications of SEM-based fixed effects panel modeling use between 3 and 10 waves of data.

In terms of across-person sample size, the N needed for adequate statistical power typically decreases as the number of time points increases, everything else being equal. Many researchers seek a sample size of 100 or more but scenarios exist were fewer individuals than this (N = 50) can be effectively analyzed. I provide a power analysis program on my website for conducting power analysis for a predictor in a traditional fixed effects regression model called *Power: FE panel data*. It does not generalize to more complex variants of SEM-based fixed effects modeling but the program can be used to gain a rough sense of power for linear fixed effects regression. If need be, you can pursue localized simulations in Mplus using the methods described in Chapter 28

Like many statistical methods tied to regression-based analytics, results can be biased by measurement error. Correctives are possible, such as using latent variables with multiple indicators. In addition, one can perform sensitivity analyses relative to measurement error as detailed in the document *Additional RET Applications of SEM-based Fixed Effects Modeling with Panel Data* on my website.

Another shortcoming to keep in mind is that although fixed effects panel methods deal reasonably well with time-invariant confounds, they are subject to bias when time-varying confounds are not statistically controlled. This is why it is important to anticipate what these time-varying confounds might be when planning an RET study and to be sure to assess them for possible model inclusion. To address the problem, some researchers include time as a fixed predictor or covariate in their models, either continuously scored or in the form of one or more dummy variables. The idea is that if one suspects there are general time trends or events that affect your panel across time, including time as a

covariate will adjust for this. Of course, there is the danger that controlling for period effects will undermine the external validity of the constructs that are of substantive interest to you because those constructs are themselves varying over time. I discuss the matter in more detail in this chapter in the context of the use of mixed modeling in RETs, where I introduce the construct of detrending.

In the final analysis, fixed effects panel models yield coefficients that are within-person focused. If your primary interest is in between-person coefficients, these models are limited. However, I find that for the analysis of RETs, both between-person and within person frameworks often are informative and, as you will see shortly, I do not hesitate to use both when trying to gain insights into intervention effects using an RET framework. Each approach taps into something slightly different and my preference is to understand intervention effects from multiple perspectives.

For useful references on SEM-based fixed effects panel modeling, see Allison (2009), Andersen, (2022), Bollen and Brand (2010), Firebaugh, Warner and Massoglia (2013), Moral-Benito et al. (2019), Murayama and Gfrörer (2024), and Quintana (2021).

SEM-BASED RANDOM EFFECTS PANEL MODELS

Equation 16.6, which I repeat here for convenience, presented the foundational equation for the fixed effects panel model:

$$Y_{it} = \mu_t + \beta X_{it} + \gamma Z_i + \alpha_i + \varepsilon_{it} \quad [16.7]$$

where μ_t is an intercept that can be different at each time point, X_{it} is a vector of predictors that change over time across the measurement occasions, Z_i is a vector of time-invariant predictors, ε_{it} is a random disturbance term, and α_i represents all differences between individuals that are stable over time (time-invariant) and not otherwise accounted for by Z_i. This equation also is foundational for an alternative to the fixed effects panel model known as a **random effects panel model** but with slightly different conceptualizations and assumptions of the equation terms. One primary difference between the two models is the treatment and conceptualization of α_i. In the fixed effects model, α_i is treated as fixed for each individual and is allowed to be freely correlated with the observed time-varying predictors in the model. In a random effects panel model, α_i is instead treated as a random variable that is normally distributed with a mean of zero, a constant variance, and a zero correlation with (or, more strongly, is statistically independent of) the observed time-varying predictors. Unlike the fixed effects panel method, random effects estimation does not discard variation across individuals. Rather, it *assumes* that the unmeasured time-invariant causes in α_i are independent of the

measured presumed causes in the model. This assumption has the advantage of producing smaller standard errors, greater statistical power, and narrower confidence intervals than the fixed effects approach, but it also comes at a cost, namely the presence of unmeasured across-individual confounds and omitted variable bias. Some methodologists refer to the independence assumption as unrealistic bordering on what they call "heroic."

On a practical level, the SEM-based programming strategy outlined in the previous sections can be amended to estimate a random effects model by eliminating the Mplus commands that specify correlations between the latent α and the exogenous variables. For the fully constrained model programmed in Table 16.5, I would delete Line 17 to estimate a random-effects panel model instead. When I did so, here is the global chi square test statistic that resulted:

```
Chi-Square Test of Model Fit

          Value                             86.926*
          Degrees of Freedom                    85
          P-Value                           0.4216
          Scaling Correction Factor         0.9879
```

The chi square statistic for the original fixed effects panel model was

```
Chi-Square Test of Model Fit

    Value                                   71.038*
          Degrees of Freedom                    70
          P-Value                           0.4429
          Scaling Correction Factor         0.9870
            for MLR
```

I used the program on my website called *Scaled chi sqr difference test* to conduct a nested chi square difference test between the two models. The scale corrected chi square difference was 15.89 with 15 degrees of freedom which yielded a p value < 0.39. This is an analog of the Hausman test often used to compare fixed and random effects models. It appears that I could use the random effects panel model in this case because there is not a meaningful difference in the global fit of the two models. However, the within-person coefficients in the fixed effects model all have small standard errors due to the large sample size, so the conclusions from a random effects model likely will not be different.

GENERALIZED LINEAR MODELS FOR PANEL DATA

Jacob Long (2019) has adapted a class of generalized linear models for the analysis of

panel data using variants of the hybrid approach of Allison (2005) that combine fixed effect panel models with across-person analyses. The approach is implemented in the R package called *panelr*, which I make available in a program on my website called *GLM panel regression*. The models are sometimes referred to as within-between or between-within models. In RETs, the method is most likely to be used to analyze the relationship between mediators and outcomes in a limited information estimation framework. Suppose as before I have two mediators, MA and MB, and a time varying covariate, COV. I measure each of these variables at the posttest and again at 3, 6, 9, and 12 month intervals during the follow-up period. Thus, I have five posttreatment time points with each yielding separate scores on Y, MA, MB and COV, i.e., they are time varying.

As noted, there are two types of predictors in panel regression. Time-varying predictors change in value from one time period to the next, and this is the case for MA, MB, and COV. Time-invariant predictors are those that are stable and do not change over time. One question I might want to address for the time-varying variables that constitute mediators and outcomes is whether changes in the mediators across time are associated with changes in the outcome across time; that is, if I work with the data within people but across time, is variation in the mediator associated with variation in the outcome across the five time points? If I regress the outcome onto the mediator within people but across time points, is the resulting regression coefficient non-zero? Panel regression using generalized linear models provides perspectives on this question by yielding estimates of such regression coefficients. These are comparable to what we get with traditional fixed effects panel regression. Suppose I find that the within-person regression coefficient for a given mediator is 0.30. This means that for every one unit the mediator changes across time for an individual, the outcome is predicted to increase 0.30 units.

Standard panel model analyses are such that they assume that any variation in this coefficient value in the sample data across different individuals is just random noise; that the within-person coefficient value of 0.30 more or less typifies everyone in the sample. If this is not the case, then specialized analytic methods are required to allow for coefficient variability across individuals, often referred to as the process of introducing **random or varying coefficients** for a given within-person predictor. The term *random* is used because the within-person coefficients across people are assumed to have a normal distribution with a mean and variance to be estimated. I describe how to implement this approach in *panelr* in the video for the *GLM panel regression* program on my website.

A second approach for estimating the effect of a predictor/mediator on the outcome emphasizes across-individual analysis rather than within-individual analysis. Simplifying a bit and as I discussed earlier, we calculate the mean of the target mediator across the 5 time points for each individual separately and we also do so for the outcome across the 5

time points for each individual separately. We then regress the mean outcome scores onto the mean mediator scores across all individuals in the study. The resulting coefficient is referred to as a between-person coefficient because it is calculated across individuals. This analysis can include multiple predictors, such as the individual mean vector of MB scores and the individual mean vector of COV scores. Importantly, unlike the within-person effects, the between-person coefficients that result are *not* robust to confounding due to time-invariant unmeasured variables, so sometimes we include in the analysis time-invariant confounds (McNeish, 2019; Long, 2024). Suppose I find that the coefficient value for the target mediator computed across individuals also is 0.30, the same value or nearly the same value as the within-person coefficient.

The fact that the two coefficients are similar reflects convergence between the two perspectives for characterizing the effect of the mediator on the outcome. An interesting feature of the *GLM panel regression* program is that it can test the statistical significance of the difference between the two types of coefficients, a test that is often called a test for **contextual effects**. If the coefficients are statistically significantly and meaningfully different from each other, then the question becomes why is this the case; why is the effect different within individuals across time as opposed to across individuals.

In the SEM-based fixed effects analysis example in this chapter, there were 1,000 individuals and 5 time periods. I re-analyzed it using the *panelr* program on my website to illustrate the above concepts. Here are the within-person results:

```
WITHIN EFFECTS:
-------------------------------------------------
             Est.    S.E.   t val.       d.f.       p
--------- ------- ------- -------- ---------- -------
ma          0.309   0.016   19.865   3997.000   0.000
mb          0.304   0.016   19.427   3997.000   0.000
cov         0.297   0.016   18.975   3997.000   0.000
-------------------------------------------------
```

For comparison purposes, here are the results from the SEM-based fixed effects analysis:

```
                                                    Two-Tailed
                    Estimate       S.E.  Est./S.E.    P-Value

 Y        ON
    MA                 0.315      0.015     20.926      0.000
    MB                 0.305      0.015     20.084      0.000
    COV                0.300      0.016     19.329      0.000
```

The results are close. Here are the results from the *GLM panel regression* program for the between-person analyses, which are not reported in SEM-based fixed effects analysis:

```
BETWEEN EFFECTS:

-------------------------------------------------------------
                     Est.    S.E.   t val.       d.f.       p
---------------- ------- ------- -------- --------- -------
(Intercept)         4.704   0.322   14.618   996.000   0.000
imean(ma)           0.395   0.061    6.470   996.000   0.000
imean(mb)           0.315   0.062    5.108   996.000   0.000
imean(cov)          0.359   0.062    5.825   996.000   0.000
-------------------------------------------------------------
```

The between-person coefficient for MA was 0.395 (±0.12, CR = 6.47, $p < 0.05$), for MB it was 0.315 (±0.12, CR = 5.11, $p < 0.05$) and for COV it was 0.359 (±0.12, CR = 5.82, $p < 0.05$). For example, for every one unit that the "stable" indicator of MA increases, the "stable" index of Y is predicted to increase .395 units. Note that these values are close to those of the corresponding predictor at the within-person level suggesting convergence from the within-person and between-person perspectives. The *GLM panel regression* program also reports significance tests for the difference between these coefficients, what it calls contextual effects. Here are the results:

```
CONTEXTUAL EFFECTS:
-------------------------------------------------------------
                     Est.    S.E.   t val.       d.f.       p
---------------- ------- ------- -------- --------- -------
(Intercept)         4.704   0.322   14.618    996.000   0.000
imean(ma)           0.086   0.063    1.360   1128.576   0.174
imean(mb)           0.012   0.064    0.181   1126.745   0.856
imean(cov)          0.062   0.064    0.980   1127.225   0.327
-------------------------------------------------------------
```

Note that the between-person coefficient in the prior analyses for the MA predictor was 0.395 and for the within-person coefficient it was 0.309, a difference of 0.086. This is the value in the above `Est.` column. The difference between the coefficients was not statistically significant nor was this the case for MB.

A limitation of the between-person estimates is that, unlike the within-person estimates, they can be biased by time-invariant confounds. For example, the between-person coefficient for MA holds constant the between-person coefficients for MB and COV but there may be time- invariant confounds, such as ZA and ZB, that need to be controlled. The program can accommodate the inclusion of time invariant covariates, as described in the video for the GLM panel regression program on my website. Here are the results for the between-person coefficients when I include these covariates:

```
BETWEEN EFFECTS:
-------------------------------------------------------------
                     Est.    S.E.   t val.       d.f.       p
---------------- ------- ------- -------- ---------- -------
(Intercept)         2.764   0.173   15.960   4991.000   0.000
imean(ma)           0.349   0.031   11.087   4991.000   0.000
imean(mb)           0.304   0.032    9.536   4991.000   0.000
imean(cov)          0.329   0.032   10.368   4991.000   0.000
za                  0.758   0.028   26.736   4991.000   0.000
zb                  0.452   0.010   47.296   4991.000   0.000
-------------------------------------------------------------
```

The coefficients for MA and MB change somewhat but not appreciably.

The *panelr* package uses forms of random intercept modeling. For continuous outcomes, the results for the within-person coefficients will generally closely approximate those from SEM-based fixed effects models but for binary outcomes this is not necessarily the case (although it usually is). The *panelr* approach is not as flexible as SEM-based fixed effects modeling but it does offer different perspectives on the estimated causal links. The package also cannot accommodate multiple indicator latent variables nor correlated disturbances whereas SEM-based fixed effects modeling can.

A strength of the *GLM panel regression* program is that you can introduce varying coefficients for one or more predictors in the within-person portion of the model. For example, suppose that instead of the coefficient for MA being the same across all individuals, I think instead that it varies from one individual to the next. The implementation of *GLM panel regression* on my website allows you to alter the model to include varying coefficients for MA or any other predictor. Here is the result when I do this for MA:

```
WITHIN EFFECTS:
-----------------------------------------------------
             Est.    S.E.   t val.       d.f.       p
--------- ------- ------- -------- ---------- -------
ma          0.309   0.016   19.892   1605.074   0.000
mb          0.304   0.016   19.460   4855.455   0.000
cov         0.297   0.016   19.007   4844.601   0.000
-----------------------------------------------------

RANDOM EFFECTS:
-----------------------------------
  Group      Parameter    Std. Dev.
---------- ------------ -----------
     id      (Intercept)    0.01913
     id           ma        0.01108
 Residual                    1.994
```

The reported coefficient value for MA is now the average of the different MA coefficients across individuals. The RANDOM EFFECTS table includes a term for MA that reports the estimated standard deviation of the coefficients across the individuals. In this case, the standard deviation is quite small (0.011), suggesting the more parsimonious model with non-varying coefficients may be preferable. You can evaluate the relative fit of the two models by comparing their AICs or BICs (which are provided on the program output), with preference being for the model with the lower AIC or BIC:

```
Nonvarying coefficient model: AIC = 21158.05, BIC = 21229.74
Varying coefficient model:    AIC = 21162.05, BIC = 21246.78
```

In this case, preference is for the non-varying coefficient model.

Parenthetically, if you use *panelr* to create R shortcut product terms in the time-varying portion of the model, the program will double demean the product term based on the work of Giesselmann and Schmidt-Catran (2020). In addition to cluster/individual mean centering the raw scores before forming the product term, *GLM panel regression* also cluster/individual mean centers the product term itself after the multiplication has occurred, a step that usually is not performed in traditional interaction analysis. For the logic of the approach, see Giesselmann and Schmidt-Catran (2020) and the video associated with the *GLM panel regression* program on my website.

GENERAL ESTIMATING EQUATIONS FOR PANEL MODELS

Another often used alternative to SEM-based fixed effects panel modeling is a method called **generalized estimating equations (GEE) panel modeling**. This approach adjusts for across-time dependencies that occur in repeated measure scenarios using a specialized estimation strategy known as **quasi-maximum likelihood**. GEE is an extension of generalized linear models with different link functions, so it can be applied to panel data with continuous outcomes (using the identity link), with binary outcomes (using either logit or probit binary links) or with count outcomes. It relies on a different statistical theory than SEM-based fixed-effects regression and the GLM based panel regression. Unlike the GLM based panel model that assumes random intercepts, GEE works with fixed intercepts; GEE estimates a single intercept for the entire population rather than a separate intercept for each individual.

In general, GEE estimates population-averaged (marginal) parameters rather than more traditional conditional parameters in most regression models. Its parameters adjust for covariates akin to the way that average marginal effects approaches do (see Chapter 5). GEE estimates reveal how the average outcome changes across the population for a unit change in the predictor taking into account the covariate distributions as a whole. By

contrast, fixed effects models estimate the effect of a predictor on an outcome holding constant the covariates at particular predictor values. Often the results of the two types of analytic methods are similar but sometimes not. The relevant statistical theory and underlying mathematics are described in Hardin and Hilbe (2012) and the adaptation of it to within-between panel designs is discussed in Long (2024) and Hoover, Shi, Burstyn and Anastos (2019).

Consider as an example a case where we have measured a continuous outcome, Y, at three points in time, yielding values of Y1, Y2, and Y3. We have two time-varying predictors, MA and MB, that also are measured at the three time points, yielding MA1, MA2 and MA3 and MB1, MB2 and MB3. We want to test if the mean Y changes over time are associated with changes in MA over time and changes in MB over time. Application of GEE panel modeling requires that the analyst specify the presumed correlation structure between the repeated measures of Y. There are four different correlation structures that are frequently evaluated, but the technique can handle any structure the analyst believes is applicable. The four structures as applied to the three outcome example are (where ρ signifies an across-person correlation between measures):

Independence

	Y1	Y2	Y3
Y1	1	0	0
Y2	0	1	0
Y3	0	0	1

,

Exchangeable

	Y1	Y2	Y3
Y1	1	ρ	ρ
Y2	ρ	1	ρ
Y3	ρ	ρ	1

First order autoregressive

	Y1	Y2	Y3
Y1	1	ρ	ρ^2
Y2	ρ	1	ρ
Y3	ρ^2	ρ	1

Unstructured

	Y1	Y2	Y3
Y1	1	ρ_{12}	ρ_{13}
Y2	ρ_{12}	1	ρ_{23}
Y3	ρ_{13}	ρ_{23}	1

The independence model assumes the Y are uncorrelated. For panel designs, this typically is not realistic. The exchangeable structure assumes the correlations between all the Ys are identical and equal to the value ρ. The first order autoregressive structure assumes the correlations follow a classic autoregressive/simplex pattern and is one that is often chosen for longitudinal panel models. The most flexible structure is the unstructured pattern in which the correlations can take any pattern. Simulations suggest

that the unstructured form coupled with the sandwich estimator is, in general, a good choice for a wide range of scenarios (Hardin & Hilbe, 2012) but it also tends to reduce statistical power and can lead to convergence issues. For this reason, some methodologists prefer the exchangeable or autoregressive structures for panel data (Hoover et al, 2019).

GEE estimation is unique because it uses the *a priori* specified correlation structure to derive estimates, standard errors and significance tests of the core parameters. This stands in contrast to the other analytic methods discussed in this chapter. Simulation studies suggest that misspecifying the correlation structure often does not create bias in model estimates but it can lower efficiency. Hoover et al. (2019) describe several scenarios where misspecification can affect bias. When specifying correlation structures, the assumption is typically made that the time intervals between successive measurements are the same. This is not mandatory, but when it is not the case, be careful to take it into account. A good practice is to estimate your model under different working dependency structures and determine the sensitivity or robustness of conclusions to variations in them.

The GEE approach does not require distributional assumptions in the way more traditional approaches do because estimation of the population-average model depends primarily on correctly specifying only a few features of the data-generating distribution (e.g., correct specification of the mean model and the working correlation structure), not the entire joint distribution. GEEs tend to provide reasonably robust estimates despite minor to moderate incorrect specifications. GEE assumes data are missing completely at random (MCAR), which is more demanding than maximum likelihood approaches.

The *panelr* program on my website has adapted the GEE method to perform within-between analyses per the GLM within-between framework but without random effects. I applied the GEE method to the example with the SEM-based fixed-effects panel model with five time points predicting Y from MA, MB and COV. Here are the within results:

```
WITHIN EFFECTS:
----------------------------------------
             Est.    S.E.   z val.       p
--------- ------- ------- -------- -------
ma          0.304   0.017   17.432   0.000
mb          0.307   0.017   17.854   0.000
cov         0.304   0.017   17.394   0.000
----------------------------------------
```

The results are reasonably close to the SEM-based fixed effects analysis. Here are the results from *panelr* for the between-person results:

```
BETWEEN EFFECTS:
-------------------------------------------------
                     Est.    S.E.   z val.       p
---------------- ------- ------- -------- -------
(Intercept)         4.699   0.329   14.267   0.000
imean(ma)           0.394   0.061    6.459   0.000
imean(mb)           0.324   0.060    5.404   0.000
imean(cov)          0.351   0.066    5.296   0.000
-------------------------------------------------
```

Here are the significance tests for the difference between the within-person and between-person coefficients (the contextual effects):

```
CONTEXTUAL EFFECTS:
-------------------------------------------------
                     Est.    S.E.   z val.       p
---------------- ------- ------- -------- -------
(Intercept)         4.699   0.329   14.267   0.000
imean(ma)           0.089   0.063    1.421   0.155
imean(mb)           0.018   0.062    0.284   0.776
imean(cov)          0.047   0.068    0.699   0.485
-------------------------------------------------
```

The difference between the coefficients was not statistically significant for any of the three predictors.

The marginal approach of GEE is sometimes framed as emphasizing population-averaged as opposed to person-specific effects. A person-specific approach yields coefficients that tell us what would happen to a person or a specific type of person defined by the values of the covariates (in a conditional sense) if that person's target predictor variable were increased by one unit. A population-averaged approach instead yields coefficients that tell us what would happen to the whole population if everyone's target predictor variable were increased by one unit. For linear panel models, the distinction between these two types of coefficients is moot in the sense that the estimation approaches yield the same results. For logistic panel models, however, person-specific coefficients typically are larger than population-averaged coefficients. The magnitude of the difference depends on the variance of α_i; the greater the departure of the α_i variance from zero, the larger the disparity is with the values of the population averaged coefficients gravitating towards zero as the variance of α_i increases.

Allison (2009) argues that one approach is not better than the other and that one's preference depends on your goals. If you want to specify how a one unit change in a predictor affects a specific type of person as defined by values on the additional covariates, then the conditional coefficient GEE approach is best. If you want instead to

gain perspectives on how an entire population would change if, say, the predictor was increased by one unit for everyone, then the GEE approach is probably best. For more details, see Allison (2009).

CONCLUDING COMMENTS ON WITHIN-BETWEEN MODELS

In sum, of the different approaches for analyzing within-person coefficients across time in a panel design, I believe that the SEM-based fixed effects panel strategy is among the best. It offers considerable flexibility for relaxing model constraints, provides useful overall model fit indices including modification indices, and allows the use of robust estimation strategies. It can adjust for measurement error and you can formally introduce time-invariant predictors into the model (see the document *Additional RET Applications of SEM-based Fixed Effects Modeling with Panel Data* on my webpage). The GLM within-between approach and the GEE within-between approach come at estimation from different vantage points but they have the advantage of providing analogous between-person coefficients that can be formally compared with the within-person coefficients. This is not the case for the SEM-based fixed effects analysis. In the final analysis, I think it is good to come at your analyses from multiple vantage points so I sometimes use two or three of these methods when analyzing within-between data depending on my goals. I have found that sometimes the SEM approach can be fickle and that using the GLM approach but requesting only a within=person analysis instead of a between-within person analysis is sometimes a good alternative.

AN RET NUMERICAL EXAMPLE

In this section, I outline a blueprint for analyzing complex RETs with multiple mediators and multiple follow-ups. Models for such designs can be complex because they can (but do not necessarily) include autoregressive effects (either first order or second order), simultaneous causal effects, reciprocal causal effects, and lagged causal effects for every pair/triad of variables. This can result in a large numbers of parameters to be estimated. In full information SEM (FISEM), unbiased estimation of the parameters often is tied to correct model specification. If parts of the model are misspecified, then the effects of this misspecification can not only affect the misspecified portions but it also can reverberate through to other parts of the model. We sometimes do not have strong theory to guide model specifications in their entirety which is why multiple mediator, multiple follow-up RET designs can be challenging. The numerical example I use has 4 posttreatment assessment periods with multiple covariates, some of which are time varying and some that are time invariant. I use only two mediators for pedagogical reasons but the strategy I

outline is easily extended to 5 or so mediators if not more. The approach is guided by the three core questions for RETs that have guided this book, namely (1) does the intervention meaningfully affect the outcome, (2) does the intervention meaningfully affect each of the presumed mediators, and (3) are each of the mediators meaningfully causally related to the outcome. Before describing my approach, I need to develop the logic of a strategy for dealing with RET complexity called **interventional effects**.

Interventional Effects

Consider a model with two presumed mediators (MA and MB) of an intervention (T) measured at three time points, *t1*, *t2*, and *t3*. Figure 16.14 presents the model I hypothesize underlies the presumed causal relationships between the variables (note: I omit covariates for presentation clarity).

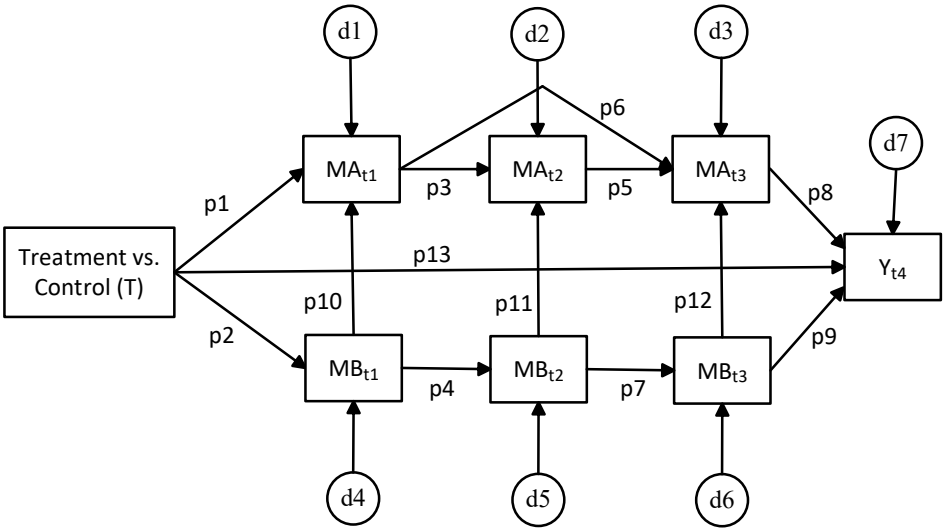


FIGURE 16.14 Multi-wave multi-mediator model

Suppose I want to determine if the intervention has an effect on the mediator MA at *t3*. In traditional FISEM, this effect is determined by algebraically combining a subset of the path coefficients from the model via the rules of path analysis (see the document on effect decomposition for Chapter 11 on my website). For our model, the formula is

$$T \rightarrow MA_{t3} = p1*p3*p5 + p1*p6 + p2*p10*p3*p5 + p2*p10*p6 + p2*p4*p11*p5 + p2*p4*p7*p12 \quad [16.8]$$

At the population level, if the model in Figure 16.14 does, in fact, reflect the true operative causal dynamics then when I combine the population level path coefficients according to Equation 16.8, the result will indeed document the impact of T on MA3. Stated another way, if I simply regress MA3 onto T at the population level and if I assume valid measures, valid random assignment, and properly distributed errors, the resulting coefficient for the regression analysis will be identical to what I would observe via Equation 16.8 under these same assumptions. However, if the causal dynamics in the model are misspecified in ways that yield incorrect path coefficients in the model or that omit important path, then Equation 16.8 mischaracterizes the effect of T on MA3.

Suppose my interest is in estimating the effect of T on MA3. I might be unsure of the complex causal structure surrounding the determinants of MA3 relative to the full set of mediators that precede or are contemporaneous with it. In this case, there is a simple work-around I can use to estimate T→MA3 (see Loh & Ren, 2022): Just regress MA3 directly onto T in a limited information sense and ignore all of the intermediatory mediational determinants and their causal relationships. This is the spirit of interventional effects, namely you estimate the effect of the randomized intervention on a target variable using stripped down regression even in the face of unknown structural causal relations among the mediators of that effect (Nguyen, Schmid & Stuart, 2021; Loh & Dongning, 2022; Loh et al., 2022; Vansteelandt & Daniel, 2017). Some researchers characterize the interventional approach as focusing on the *marginal means* of MA3 for T (Loh et al., 2022). Of course, if there are confounds then you must control for those confounds via the use of covariates, but given random assignment to T, confounds typically will be irrelevant. I make use of this interventional strategy in the numerical example.

The RET

The example RET focuses on an intervention to increase parent intentions to obtain a specific vaccine for their child when the child reaches age 10. The intervention emphasized two facets of the intervention, (a) its effectiveness in preventing the disease the vaccine targets, and (b) the likelihood that the child would contract the disease and the consequences that could follow if they did not get the vaccine. I refer to these presumed mediators of the intervention as *vaccine effectiveness* (which I will call MA) and *perceived susceptibility* (which I call MB). The intention to obtain the vaccine when the child turns 10 years of age (Y) and the two mediators were measured as the average item response to multi-item scales in which each item was responded to on a disagree-agree scale from strongly disagree (-5), to 0 (neither agree nor disagree) to +5 (strongly agree). Total scores thus could range from -5 to +5. In general, the standard deviations of each measure were approximately 1.0. Measures were obtained at baseline, immediately

after the intervention and then at 3 month, 6 month, and 12 month follow-ups. There were two time-invariant covariates, ZA and ZB, and one time-varying covariate, COV.

For the working causal model I assume that MA and MB each have linear and independent effects on Y either across individuals at a given time point or across time as MA and MB vary from one time point to the next. I also assume that perceptions of effectiveness and perceptions of susceptibility at a given point in time have near instantaneous impact on a parent's intention/decision to obtain the vaccination at that time. That is, if I convince parents that the vaccine is effective and that their child is highly susceptible to the disease the vaccine targets if they do not get the vaccine, parents will decide on the spot to pursue the vaccine at age 10, everything else being equal. Another assumption I make is that MA and MB may be correlated but there is no causal link between them. I use *t0*, *t1*, *t2*, *t3* and *t4* to refer to the different time points studied.

Including the covariates and the various time-invariant variables, the model contains 22 variables, which is on the small side compared to real-world RETs that usually include more mediators, more covariates, and, perhaps, multiple outcome variables. To apply FISEM, I must specify all of the presumed causal relationships between the 22 variables including first and/or second order autoregressive effects, contemporaneous effects, reciprocal effects, first and/or second order lagged effects, and the presumed relationships among the disturbances for the 15 or so endogenous variables. FISEM estimates of the total effect of the intervention on the outcome(s), the effects of the intervention on each of the mediators, and the effects of the mediators on the outcome are all dependent on a correctly specified model. Suppose I have concerns about applying FISEM given the model complexity and the state of current substantive theory to guide my choices. Stated another way, am I really all that confident that there is no specification error in the model I formulate for purposes of FISEM testing and for answering the fundamental questions of an RET? Given the many possible alternative model structures, I might decide to shift to a more compartmentalized LISEM approach.

Total Effect of the Treatment on the Outcome

The estimated overall effect of the treatment condition on the outcome requires a between-person perspective because the intervention vs. control manipulation is inherently and strictly across individuals in nature. i.e. there is no across-time variation in it. Figure 16.15 presents the model I test to evaluate the relevant total effects using interventional logic. The intervention is presumed to influence the outcome at the immediate posttest, the 3 month follow-up, the 6 month follow-up and the 12 month follow-up as signified by p_1, p_2 , p_3 and p_4. The baseline outcome serves as a covariate for the outcome at each time point as does the baseline COV covariate. These variables are

included in the analysis to augment statistical power for the path coefficients of interest. For the disturbances, the dashed line is a short hand way of signifying that all of the disturbances are to be correlated with each other rather than drawing all the traditional pairwise curved arrows. This reduces clutter in the diagram. The disturbance correlations accommodate the associations between the Ys at the different time points over and beyond the common causal treatment condition determinant and the two covariates. The model underscores there is no single "total effect" for the intervention. Rather, there are four time points for which I want to document the intervention effect.

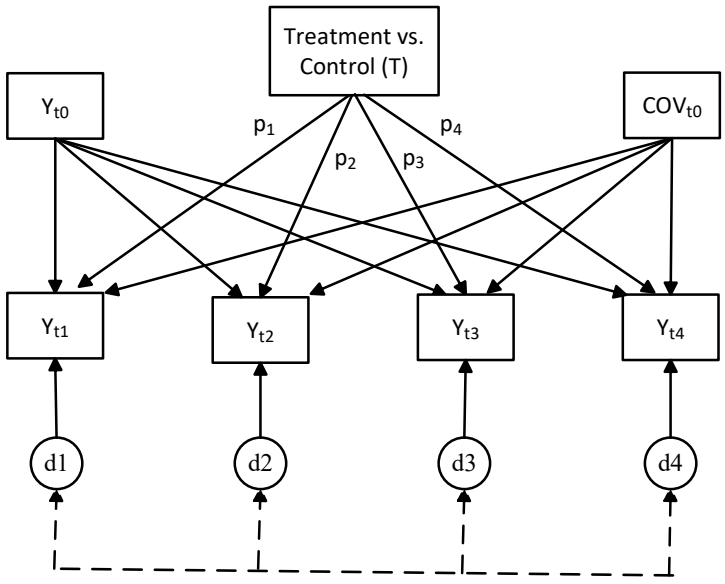


FIGURE 16.15 Evaluation of the total effect of the intervention

Table 16.7 presents the relevant Mplus syntax.

Table 16.7: Mplus Syntax for Total Effect Analysis

```
1.  TITLE: Evaluation of total effects
2.  DATA: FILE = FEmainM.dat ;
3.  DEFINE:
4.   CENTER cov0 y0 (GRAND) ;
5.  VARIABLE:
6.    NAMES ARE id za zb cov0 cov1 cov2 cov3 cov4
7.      ma0 ma1 ma2 ma3 ma4 mb0 mb1 mb2 mb3 mb4
8.      y0 y1 y2 y3 y4 treat covmean mamean mbmean ymean;
9.  USEVARIABLES ARE treat cov0 y0 y1 y2 y3 y4 ;
10.  ANALYSIS: ESTIMATOR=MLR ;
```

```
11.  MODEL:
12.  y1 y2 y3 y4 ON treat (p1-p4) ;
13.  y1 y2 y3 y4 ON y0 cov0 ;
14.  y1 y2 y3 y4 WITH y1 y2 y3 y4 ;
15.  OUTPUT: SAMP STDYX MOD(4) CINTERVAL RESIDUAL TECH4  ;
```

All of the syntax should be self-explanatory. Note that I mean center the two covariates on Lines 3 and 4 to make the intercepts for Y1 to Y4 more interpretable, namely they will now reflect the covariate adjusted Y means for the control group at each time point. Line 12 regresses each posttest Y variable onto the treatment condition and also labels the four resulting coefficients p1, p2, p3 and p4 for later use in supplementary analyses. Line 13 adds to the equations the two covariates. Line 14 correlates all the disturbances to accommodate presumed correlations between the different Y. The model is just identified which assures perfect fit. Inspection of the model indices of fit are moot. Here is the output for the effects of the intervention on Y at each time point:

		Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
Y1	ON				
TREAT		1.932	0.065	29.577	0.000
Y0		0.076	0.031	2.407	0.016
COV0		-0.008	0.034	-0.253	0.800
Y2	ON				
TREAT		1.974	0.066	30.027	0.000
Y0		0.094	0.032	2.943	0.003
COV0		-0.023	0.034	-0.664	0.507
Y3	ON				
TREAT		1.022	0.065	15.775	0.000
Y0		0.121	0.032	3.739	0.000
COV0		-0.051	0.036	-1.443	0.149
Y4	ON				
TREAT		0.113	0.067	1.697	0.090
Y0		0.083	0.030	2.753	0.006
COV0		0.017	0.035	0.492	0.623

The predicted mean difference between the intervention and control groups at the immediate posttest was 1.93 (±0.13, CR = 29.58, $p < 0.05$), which was statistically significant. Suppose prior to the study it was decided the standard for a meaningful difference was 0.50 for each of the four contrasts. Because the lower limit of the 95% confidence interval for Y1 was larger than this effect size standard (1.80 versus 0.50), I conclude that the intervention had a meaningful effect on the outcome. Examination of

the other coefficients (p2, p3 and p4) shows meaningful effects at the 3 and 6 month time points but not for 12 months.

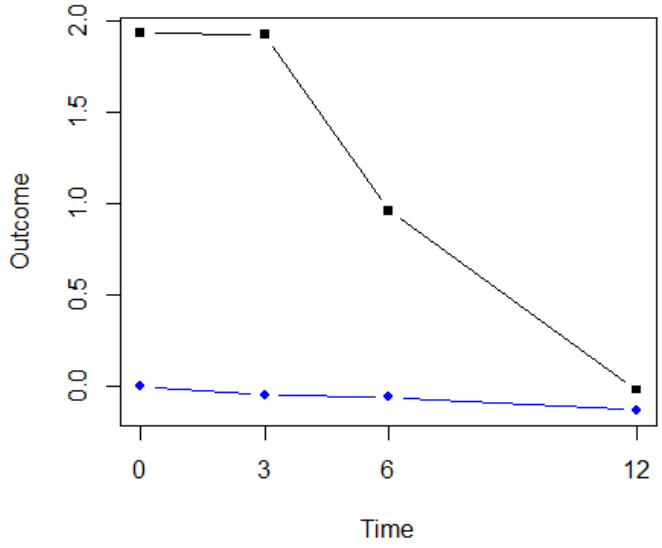
Here are the relevant intercepts for Y1 through Y4:

```
                                                    Two-Tailed
                   Estimate       S.E.  Est./S.E.    P-Value
Intercepts
    Y1               -0.003      0.046     -0.071      0.943
    Y2               -0.049      0.047     -1.032      0.302
    Y3               -0.060      0.047     -1.284      0.199
    Y4               -0.131      0.048     -2.706      0.007
```

The intercepts are the covariate adjusted estimated Y means for the control group. The margins of error for the means are, roughly, double the estimated standard error associated with each mean (`S.E.` on the output), or about ±0.09. You can obtain the corresponding intervention group means by reverse coding the treatment dummy variable by adding the command `treat = abs(treat - 1);` just after Line 4 in Table 16.7. Then re-run the program. Here are the estimated intercepts that result in this new run, each of which represents the covariate adjusted means for the intervention group:

```
                                                    Two-Tailed
                   Estimate       S.E.  Est./S.E.    P-Value
Intercepts
    Y1               1.929      0.046     41.639      0.000
    Y2               1.925      0.046     42.002      0.000
    Y3               0.962      0.045     21.409      0.000
    Y4              -0.018      0.046     -0.381      0.703
```

I used the program on my website called *Temporal line plot* to plot the means for the two groups, as follows:



The black line is the intervention group and the blue line is the control group. The distance between the lines at a given time point represents the effect of the intervention at that time point with larger distances reflecting larger effects. The means for the control group are relatively constant across time. The means for the intervention group tend to decline across time but non-linearly; there is negligible decline from the immediate posttest to month 3 after which there is a decline for the later time points.

Using the syntax in Table 16.7, I can formally test if the intervention versus control differences at the various time points are statistically significantly different from one another by adding the following model constraint commands to the syntax just before Line 15 and removing the `MOD(4)` option from the `OUPTUT` line :

```
MODEL CONSTRAINT:
 NEW(diff1 diff2 diff3 diff4 diff5 diff6);
 diff1 = p1-p2;
 diff2 = p1-p3;
 diff3 = p1-p4;
 diff4 = p2-p3;
 diff5 = p2-p4;
 diff6 = p3-p4;
```

The `MODEL CONSTRAINT` command should be familiar to you from previous chapters. To use it, you must remove `MOD(4)` from the `OUTPUT` line. I perform six pairwise contrasts labeled diff1 through diff6. The contrasts focus on the path coefficients labeled p1 through p4 in the original syntax. Here are the results:

```
MODEL RESULTS

                                                    Two-Tailed
                  Estimate       S.E.  Est./S.E.    P-Value
New/Additional Parameters
    DIFF1           -0.041      0.089     -0.466      0.641
    DIFF2            0.911      0.089     10.187      0.000
    DIFF3            1.819      0.090     20.140      0.000
    DIFF4            0.952      0.088     10.779      0.000
    DIFF5            1.861      0.088     21.085      0.000
    DIFF6            0.909      0.092      9.855      0.000
```

All of the results of the contrasts were statistically significant except for the contrast for the immediate posttest group difference vs. the 3 month group difference. Note that when evaluating decay I made sure to use contrasts that include both the intervention and control conditions. This is important to maintain the spirit of the randomized design.

The analytic strategy described in this section is a form of multivariate analysis of covariance (Bray & Maxwell, 1985; Raush, Maxwell & Kelley, 2003). The use of Mplus to accomplish it allows me to use robust estimation (you also can use bootstrapping in addition to MLR) and state of the art missing data methods. In the analysis, I used the baseline Y_{t0} as a covariate for the contrasts on the separate posttreatment Y variables following the recommendations of Raush et al. (2003; see their Model 1 for PPF designs).

Parenthetically, if the outcome is binary, then it is not straightforward to apply the multivariate model of Figure 16.15. One strategy is to use the modified linear probability model (MLPM) to estimate covariate adjusted average marginal effects in the form of risk differences of the predictors (see Chapters 5 and 12). You also can evaluate the contrasts for the `MODEL CONSTRAINTS` command in the same way as above, although you probably should conduct a second sensitivity analysis using bootstrapping to estimate standard errors, p values, and confidence intervals of the different contrasts. This involves changing Line 10 in Table 16.7 to read

```
ANALYSIS: ESTIMATOR=ML ; BOOTSTRAP = 2000 ;
```

and on the `OUTPUT` line, removing `MOD(4)` and changing `CINTERVAL` to `CINTERVAL(BOOTSTRAP)` after adding the `MODEL CONSTRAST` statements. Note that RCTs with binary outcomes often do not have a baseline measure of the outcome in which case you would just omit it from the syntax in Table 16.7. Indeed, you may choose not to include any baseline covariates, which is not problematic (other than potentially reducing statistical power) given random assignment to the treatment condition.

If you are uncomfortable with the MLPM, you can instead estimate the four equations, one for each time point, separately rather than multivariately using probit or

logit modeling in a LISEM framework per the strategies described in Chapter 12.

Effect of the Treatment on the Mediators

To evaluate the effect of the treatment condition on the mediators I use the identical analytic strategy as that for the treatment effect on the outcome but substitute the measures of each mediator for the outcome measures. The syntax I use for the analysis of mediator MA is in Table 16.8:

Table 16.8: Mplus Syntax for Treatment Effect on Mediator MA

```
1.  TITLE: Evaluation of mediator MA
2.  DATA: FILE = FEmainM.dat ;
3.  DEFINE:
4.   CENTER cov0 ma0 (GRAND) ;
5.  VARIABLE:
6.    NAMES ARE id za zb cov0 cov1 cov2 cov3 cov4
7.      ma0 ma1 ma2 ma3 ma4 mb0 mb1 mb2 mb3 mb4
8.      y0 y1 y2 y3 y4 treat covmean mamean mbmean ymean;
9.  USEVARIABLES ARE treat cov0 ma0 ma1 ma2 ma3 ma4 ;
10.  ANALYSIS: ESTIMATOR=MLR ;
11.  MODEL:
12.  ma1 ma2 ma3 ma4 ON treat (p1-p4) ;
13.  ma1 ma2 ma3 ma4 ON ma0 cov0 ;
14.  ma1 ma2 ma3 ma4 WITH ma1 ma2 ma3 ma4 ;
15.  OUTPUT: SAMP STDYX MOD(4) CINTERVAL RESIDUAL TECH4  ;
```

Here are the core results:

```
                                                    Two-Tailed
                   Estimate       S.E.  Est./S.E.    P-Value

MA1      ON
   TREAT              1.885      0.063     30.072      0.000
   MA0                0.014      0.031      0.448      0.654
   COV0               0.010      0.032      0.306      0.760
MA2      ON
   TREAT              2.018      0.063     32.254      0.000
   MA0                0.006      0.031      0.200      0.841
   COV0              -0.029      0.032     -0.897      0.370

MA3      ON
   TREAT              1.002      0.064     15.707      0.000
   MA0               -0.019      0.031     -0.612      0.541
   COV0               0.008      0.032      0.240      0.811
```

```
 MA4        ON
    TREAT              0.129      0.064      2.020      0.043
    MA0                0.007      0.030      0.249      0.804
    COV0              -0.003      0.031     -0.100      0.921
```

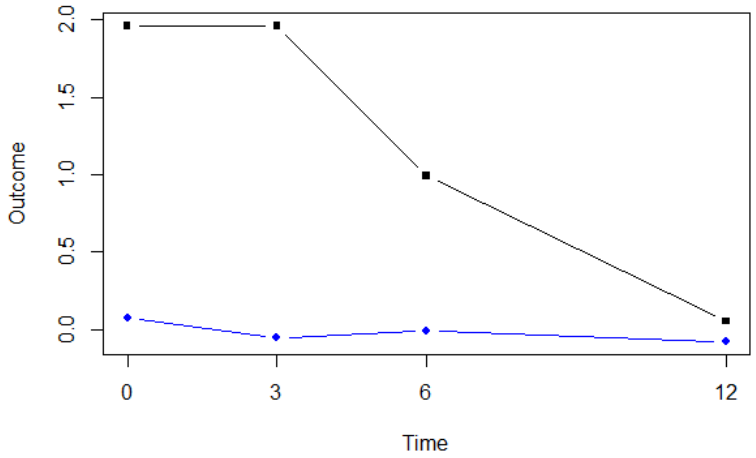
Here are the intercepts that reflect the covariate adjusted means for the control group:

```
                                                    Two-Tailed
                   Estimate       S.E.  Est./S.E.    P-Value
Intercepts
    MA1                0.075      0.045      1.663      0.096
    MA2               -0.056      0.044     -1.260      0.208
    MA3               -0.009      0.045     -0.197      0.844
    MA4               -0.077      0.045     -1.712      0.087
```

Here are the intercepts for the reverse scored treatment dummy variable analysis, which yields the estimated covariate adjusted means for the intervention group:

```
                                                    Two-Tailed
                   Estimate       S.E.  Est./S.E.    P-Value
Intercepts
    MA1                1.960      0.043     45.072      0.000
    MA2                1.962      0.044     44.332      0.000
    MA3                0.993      0.045     22.164      0.000
    MA4                0.052      0.045      1.145      0.252
```

The line plot of the group means is as follows:



It shows the same pattern as that for the outcome, Y. When I added the `MODEL CONSTRAINTS` commands for the mean difference contrasts, they yielded the same significance patterns as for the outcome.

When I analyzed MB using the same approach, all of the results paralleled those of MA and also of Y.

In sum, the intervention meaningfully affected both of the presumed mediators at the immediate posttest and at 3 months after the treatment but this effect weakened by six months and seemed to dissipate fully by 12 months.

Effect of the Mediators on the Outcome

Because the mediators and the outcomes were measured at each time point, I can approach analyses of the effects of the mediators on the outcome from either a within-individual perspective, an across-individual perspective, or both. As noted, it is reasonable for the current example to assume the operative time lag for the effect of the mediator to translate into an outcome effect is short or near instantaneous. Thus, working with a model that assumes contemporaneous effects is appropriate. I also assume the two mediators have unidirectional linear relationships with the outcome and do not impact one another, i.e., that can use a linear model to describe their relationship to the outcome.

Within-Person Analysis

My primary focus for the within-person analysis is to estimate the path coefficients that describes how changes in the mediators across time are associated with changes in the outcome across time after controlling for all time-invariant confounds, either measured or unmeasured as well as the time-varying confound COV. The resulting path coefficients provide me with perspectives on how a one unit change in a given mediator translates into change on the outcome across time. The model closely follows the logic of Figure 16.13 and the syntax of Table 16.5 that I presented earlier.

An issue I must consider is how to treat the baseline measures MA_{t0}, MB_{t0}, and Y_{t0} in the model. These measures are taken prior to the intervention, with 50% of the sample (namely those in the intervention condition) then exposed to activities designed to induce non-trivial change in the constructs. Individuals in the control group, although not exposed to an intervention, also likely experiences natural fluctuations in the post-baseline MA, MB, and Y constructs. The baseline measures no longer serve the function of across-individual confound control because all such confounds are already controlled by virtue of the use of SEM-based fixed effect modeling. Despite this, one also might argue that within-individual fluctuations from the baseline to the immediate posttest can inform how changes in the mediators are associated with changes in the outcome and

hence, should be a part of the model. In the analyses I report, I included them but some contexts might lead to a different decision. The relevant Mplus syntax is in Table 16.9.

Table 16.9: Mplus Syntax for Effect of Mediators on Outcome

```
1.  TITLE: Evaluation of mediators on outcome
2.  DATA: FILE = FEmainM.dat;
3.  VARIABLE:
4.    NAMES ARE id za zb cov0 cov1 cov2 cov3 cov4 ma0 ma1 ma2
5.    ma3 ma4 mb0 mb1 mb2 mb3 mb4 y0 y1 y2 y3 y4 treat covmean
6.    mamean mbmean ymean ;
7.  USEVARIABLES ARE ma0 ma1 ma2 ma3 ma4 mb0 mb1 mb2 mb3 mb4
8.    cov0 cov1 cov2 cov3 cov4 y0 y1 y2 y3 y4 ;
9.  ANALYSIS: ESTIMATOR=MLR ;
10. MODEL:
11.  alpha BY y0@1 y1@1 y2@1 y3@1 y4@1; !define latent time-invariant vars
12.  y0 ON ma0 mb0 cov0 (p1 p2 b1) ; !regress t0 y onto t0 preds
13.  y1 ON ma1 mb1 cov1 (p1 p2 b1) ; !regress t1 y onto t1 preds
14.  y2 ON ma2 mb2 cov2 (p1 p2 b1) ; !regress t2 y onto t2 preds
15.  y3 ON ma3 mb3 cov3 (p1 p2 b1) ; !regress t3 y onto t3 preds
16.  y4 ON ma4 mb4 cov4 (p1 p2 b1) ; !regress t4 y onto t4 preds
17.  y0 y1 y2 y3 y4 (evar);          !estimate disturbance variances
18.  alpha WITH ma0-ma4 mb0-mb4 cov0-cov4 ; !correlate latent var with all Xs
19.  OUTPUT: Samp StdYX MOD(4) Residual Tech4  ;
```

All of the syntax should be familiar. Upon execution, the indices for global fit all suggested satisfactory fit. The chi square test of perfect model fit was statistically non-significant (chi square = 64.81 with 70 degrees of freedom, $p < 0.66$), which is consistent with the proposition that the data are reasonably in accord with the model. The RMSEA was less than 0.001 (90% confidence intervals of 0.00 to 0.016), the p value for close fit was non-significant ($p < 1.00$), the comparative fit index was 1.00, and the standardized RMR was 0.01. There were no meaningful modification indices above 4.0 nor were any of the standardized residuals in the form of z tests statistically significant.

Here are the core results for the path coefficients:

```
MODEL RESULTS

                                                    Two-Tailed
                  Estimate       S.E.  Est./S.E.    P-Value

 Y0       ON
    MA0               0.495      0.010     50.114      0.000
    MB0               0.492      0.010     49.778      0.000
    COV0              0.218      0.011     19.687      0.000
```

```
Y1         ON
   MA1                 0.495      0.010     50.114      0.000
   MB1                 0.492      0.010     49.778      0.000
   COV1                0.218      0.011     19.687      0.000

Y2         ON
   MA2                 0.495      0.010     50.114      0.000
   MB2                 0.492      0.010     49.778      0.000
   COV2                0.218      0.011     19.687      0.000

Y3         ON
   MA3                 0.495      0.010     50.114      0.000
   MB3                 0.492      0.010     49.778      0.000
   COV3                0.218      0.011     19.687      0.000

Y4         ON
   MA4                 0.495      0.010     50.114      0.000
   MB4                 0.492      0.010     49.778      0.000
   COV4                0.218      0.011     19.687      0.000
```

The coefficient for the effect of the mediator MA on Y was 0.50 (±0.02, CR = 50.11, $p < 0.05$) and for MB it was 0.49 (±0.02, CR = 49.78, $p < 0.05$). For each mediator, for every one unit that it increases across time for people, the intention to obtain a vaccine increases by approximately 0.50 units. Suppose prior to the study it was decided that a population coefficient of 0.20 or greater would be meaningful. Because the lower limits of the 95% confidence interval for both coefficients were greater than this standard, the effects are declared meaningful.

Some researchers end the exploration of mediator effects on the outcome here because they feel the fixed effects analyses that control for all measured and unmeasured time invariant confounds as well as the measured time-varying confounds (in this case, COV) are on stronger footing than the across-person analyses that may have unmeasured time invariant confounds. Others prefer to explore across-person perspectives in order to gain additional perspectives on matters. For the between-person case, the focus is not on explaining how changes in mediators affect changes in outcomes over time within individuals. Rather the focus is on explaining what makes some people have higher outcome scores than other people with the likely culprits being elevated scores on the mediators.

Across-Person Analysis

There are different forms that the across-person analyses can take. One informative strategy is to regress the outcome onto the mediators, the treatment condition, and the baseline Y and baseline COV measures vis-à-vis traditional across-person linear

regression using Mplus at each of the posttreatment time points, separately. Table 16.10 presents the syntax for the first regression analysis and Table 16.11 presents the estimated path coefficients for the two mediators and the treatment condition direct effect for all the analyses. The baseline Y_{t0} is used in each regression analysis to control (imperfectly) for unmeasured time-invariant confounds, as I have done in previous chapters, as is COV_{t0}. The direct effect of the treatment cannot be evaluated in the prior within-person analyses because variation in the treatment condition only occurs at the across-person level. Each model is just-identified so matters of model fit are moot.

Table 16.10. Mplus Syntax for Linear Regression at a Given Time Point

```
1.  TITLE: Regression at immediate posttest
2.  DATA: FILE = FEmainM.dat;
3.  VARIABLE:
4.    NAMES ARE id za zb cov0 cov1 cov2 cov3 cov4 ma0 ma1 ma2
5.    ma3 ma4 mb0 mb1 mb2 mb3 mb4 y0 y1 y2 y3 y4 treat covmean
6.    mamean mbmean ymean ;
7.  USEVARIABLES ARE ma1 mb1 cov1 cov0 y0 treat ;
9.  ANALYSIS: ESTIMATOR=MLR ;
10. MODEL:
11.  y4 ON ma4 mb4 treat cov0 y0  ;
12.  y0 ON ma0 mb0 cov0 (p1 p2 b1) ; !regress t0 y onto t0 preds
13.  y1 ON ma1 mb1 cov1 (p1 p2 b1) ; !regress t1 y onto t1 preds
14.  y2 ON ma2 mb2 cov2 (p1 p2 b1) ; !regress t2 y onto t2 preds
15.  y3 ON ma3 mb3 cov3 (p1 p2 b1) ; !regress t3 y onto t3 preds
16.  y4 ON ma4 mb4 cov4 (p1 p2 b1) ; !regress t4 y onto t4 preds
17.  y0 y1 y2 y3 y4 (evar);          !64stimate disturbance variances
18.  alpha WITH ma0-ma4 mb0-mb4 cov0-cov4 ; !correlate latent var with all Xs
19.  OUTPUT: Samp StdYX MOD(4) Residual Tech4  ;
```

Table 16.11. Results of Across-Person Regressions

Outcome	MA Coeff	MB Coeff	Treat Coeff
Y1	0.50* ±0.05	0.48* ±0.05	0.05 ±0.17
Y2	0.52* ±0.05	0.52* ±0.05	-0.07 ±0.17
Y3	0.48* ±0.05	0.47* ±0.05	0.02 ±0.13
Y4	0.48* ±0.05	0.51* ±0.05	0.03 ±0.10

(notes: * $p < 0.05$; numbers after ± are the margins of error based on the 95% confidence intervals)

In Table 16.11, there is little evidence for a direct effect of the treatment condition on the outcome independent of the mediators at any of the time points. The coefficients for MA are quite similar at each of the time points and this also is true for MB. The across-person coefficients also are close in value to what I found in the within-person analyses. All of the MA and MB coefficients are statistically significant ($p < 0.05$). The conclusions I make about MA and MB in the across-person analyses replicate the conclusions I make in the within-person analyses.

It is important to note that the across-person regression strategy I used here is appropriate when there are no causal relationships among the mediators, when the mediators combine additively to impact the outcome, and when there are no meaningful causal autoregressive effects for prior Y on the current Y reflecting meaningful inertia. If the working structural model linking mediators to Y is more complex than this, then the across-person analyses can still be performed in the spirit of Table 16.10 but instead of a simple linear regression at each time point, the more complicated structural submodel at each time point is analyzed with Mplus.

In sum, both the within-person and across person analyses are consistent with meaningful effects of the two mediators on the outcome.

Omnibus Mediation Analysis

As noted throughout this book, I do not prioritize the omnibus indirect effect analysis. One can use the joint significance test to evaluate the statistical significance of the omnibus indirect effects by determining if each link in a given mediational chain is statistically significant, i.e., that there are no "broken" links in a chain. Granted sometimes this involves mixing within-person and across-person tests but doing so likely is reasonable in most cases. In the current instance, the omnibus test for MA and the omnibus test for MB were both statistically significant via joint significance.

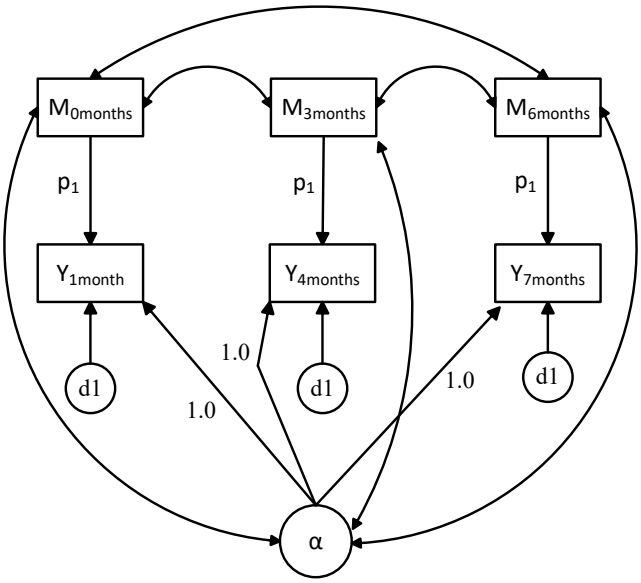
Concluding Comments for the Numerical Example

On a big picture level, the results suggest that the intervention had a meaningful impact on the outcome at the immediate and 3 month posttests. These effects diminished but remained meaningful at the 6 month posttest. By the 12 month posttest, the effect of the intervention had dissipated, suggesting the need for a booster intervention at that time. The effects of the intervention on the two mediators followed this same pattern and results suggested that mediational change meaningfully impacted outcome change.

The numerical example illustrated the SEM-based fixed effects approach using a common but simplistic model structure, namely a model with contemporaneous effects of the mediators on Y for the within-person analyses coupled with the absence of

autoregressive and cross-lagged effects for the various Y. To be sure, the model allowed for autocorrelations across time for the different Ys but it assumed that the source of these correlations was the common cause of (unmeasured) time-invariant variables and/or the correlations between the predictors covariates across time. If I instead believed there were meaningful within-person autoregressive or lagged causal effects relative to the time periods used after controlling for all time-invariant variables, the model would need to be modified to accommodate these effects. In the document on my website *Additional RET Applications of SEM-based Fixed Effects Modeling with Panel Data*, I describe a range of alternative models that you might explore in the SEM-based fixed effects framework. I also extend the model to the analysis of binary, ordinal, and count outcomes. This document is an important read.

For the numerical example, I justified the contemporaneous structure of the SEM-based fixed effects model because of the short time interval governing how long it takes for the intervention to affect the mediators and for the mediators, in turn, to affect the outcome. Interestingly, the "contemporaneous" structure of the model can be invoked in scenarios where the causal translation lags are lengthy. Consider the case of a single mediator, M, and a single outcome Y. M is measured at three points in time with 3 month intervals between the measures, yielding $M_{0months}$, $M_{3months}$, and $M_{6months}$. Y also is measured at three time points but the measures are obtained one month after each mediator to give the mediator time to translate into Y. Thus, Y is Y_{1month}, $Y_{4months}$, and $Y_{7months}$. The SEM-based fixed effect model can be diagrammed "contemporaneously" analogous to Figure 16.13 as follows:



In this figure, I use a common label for the path coefficients (p_1) to reflect a coefficient equality constraint and also for the disturbance terms (d1) to reflect a disturbance variance equality constraint. Note that the "contemporaneous" structure of the model is present so that I can apply the same analytic strategy used for the numerical example in this chapter. I just need to keep in mind the variable content and time frames when interpreting the results. In this sense, the "contemporaneous" effect causal model might be more general than one might think initially.

LATENT GROWTH CURVE MODELING

Latent growth curve modeling is another analytic technique for analyzing RET data with multiple follow-ups. Traditional growth curve analysis is unique in that it models trajectories of change in a variable across time, typically with a focus on isolating possible determinants of individual differences in the trajectories. Consider four elementary grade students (S1, S2, S3, and S4) whose reading scores are tracked over four six month intervals, per Figure 16.16.

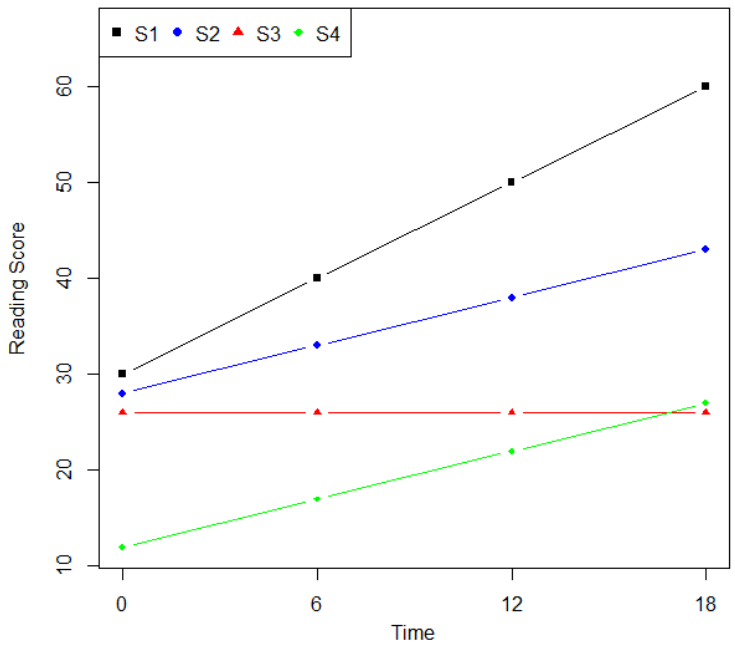


FIGURE 16.16 Reading trajectories of four students

There are interesting features of the different trajectories. First, note that all of them are linear. Such perfect linearity is rare in practice but often the trajectories approximate linear functions closely enough that they can be treated as being functionally linear. Deviations from linearity may occur for a given individual but these deviations typically are conceptualized as random disturbances from the true underlying linear function.

Second, for each individual, I can compute a value for his or her intercept (his or her reading score when Time = 0) and his or her slope, with the latter constituting a "score" representing his or her trajectory. In this case, S1 has an intercept of 30 and a slope of 10; S2 has an intercept of 28 and a slope of 15; S3 has an intercept of 26 and a slope of 0; S4 has an intercept of 12 and a slope of 5. Note that the larger the slope, the steeper is the student's trajectory reflecting more dramatic improvement in reading across time. Students S2 and S4 show equivalent upward trajectories (slopes = 5) even though they start from different initial reading levels as reflected by their different intercepts (the intercept for S2 is 28 and for S4 it is 12). S1 shows considerable improvement in reading across time (slope = 10), while S3 shows no improvement whatsoever (slope = 0).

A researcher might want to isolate possible causes of these individual differences in trajectories. Perhaps the type of school a student is in, public versus private, can explain the differences in trajectories. Or, perhaps SES can explain some of the across-person variation in the trajectories. Growth modeling is designed to address such questions.

In the context of RCTs, it might be hypothesized that participation in an intervention impacts the magnitude of the outcome trajectory of individuals. Individuals in the control condition will exhibit trajectories that reflect the absence of the intervention whereas individuals in the intervention condition exhibit trajectories that result after exposure to the intervention. Typically, researchers calculate the average trajectory (slope) in each of the two groups (intervention versus control) and then compare these averages to determine if the intervention has impacted outcome trajectories. They also might measure the trajectories of presumed mediators of treatment-outcome effects and then determine if there is an association between the trajectories observed for the mediators with the trajectories observed for the outcomes. If so, the researcher might conclude the association is consistent with a causal effect of M on Y.

The term "growth modeling" is a bit of a misnomer because it implies that the phenomena being studied must be growth-like in nature. However, the analytic techniques can be applied to any outcome in which trajectories across time are the focus. Also, the methods are most appropriate when interest is in describing the omnibus shape of the trajectories and correlates of it. Indeed, some argue that the primary goal of growth modeling is to summarize a set of repeated measures with a smoothed underlying function that parsimoniously represents how a variable changes over time. If your interest

instead is describing changes that occur at specific time points by contrasting treatment and control groups at those time points, then you might consider using the methods I described above in the RET numerical example that included SEM-based fixed effects analyses. You also might find that applying both approaches yields complementary insights into RET dynamics. The key point here is that growth curve models are typically applied when interest is in longitudinal trajectories.

Growth curve modeling has been dominated by two statistical traditions, one that uses classic multilevel modeling and the other that uses structural equation modeling. The former relies on statistical packages like HLM and the R packages *lme4* and *nlme*. The latter relies on Mplus, lavaan, and other SEM software. I prefer the SEM-based approaches because they have more flexibility when evaluating mediation dynamics and they can more readily deal with measurement error and multiple indicators.

The Traditional Latent Growth Curve Model

In this section, I describe the traditional SEM-based linear growth model for a single variable, Y. This sets the stage for the more complex application of the approach to RETs that I describe later. Assume Y is measured at 4 points in time with the same time lags between the measures, say, 6 months. Equal time lags are not required but I will assume such is the case for pedagogical reasons. Y0 is the initial measure of Y, Y1 is the measure of Y after 1 lag, Y2 is the measure of Y after 2 lags, and Y3 is the measure of Y after 3 lags. Assume Y is measured on a metric from -5 to 5 with standard deviations near 1.0. Our task is to map how Y changes as a function of time for each individual in the data, that is to represent the best fitting intercept and slope for each individual in the equation

$$Y = a + b\ Time$$

The intercept will be the value of Y when Time = 0, i.e., the Y value at the first time point. The slope estimates how many units Y changes given a one lag unit increase in time, i.e., when the lag goes from 0 to 1, from 1 to 2, and from 2 to 3. Of course, this equation will not fit each individual's data perfectly; there will be an error or disturbance term in the above equation. However, the values we use for *a* and *b* are values that minimize the sum of the squared errors. Figure 16.17 presents the relevant SEM model.

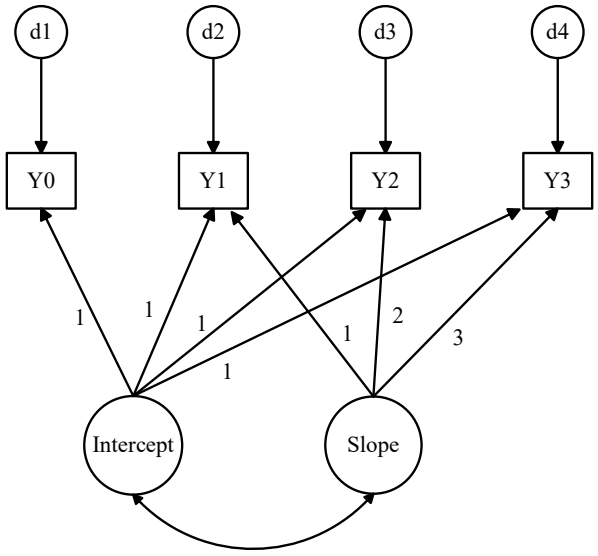


FIGURE 16.17 Traditional linear growth curve

There are two latent variables underlying the four Y measures but these latent variables are defined differently than traditional latent variables. Note that all of the factor loadings from each latent variable to each Y variable are fixed at a predetermined value. These values are carefully chosen *a priori* by the researcher so that the latent variables take on specific meanings, namely so that they reflect certain properties of the intercepts and slopes of the aforementioned linear equation. Specifically, the mean of the latent variable called 'Intercept' will equal the mean intercept calculated across the separate intercepts for each individual in the sample. The mean of the latent variable called 'Slope' estimates the average slope calculated across the separate slopes for each individual. Similarly, the standard deviation of 'Intercept' reflects the variability in the individual intercepts and the standard deviation of 'Slope' reflects the variability in the individual slopes. The latter often is of interest because it reflects how much variability there is in the linear trajectories across individuals.

The underlying mathematics for the values at which to fix the loadings are described in Bollen and Curran (2006) and Preacher et al. (2008). Note that the 'Intercept' latent variable loadings are always fixed to the value of 1.0 for each Y variable. The loadings for the 'Slope' latent variable are fixed at the respective number of lags for each Y, namely 0, 1, 2, and 3 for Y0, Y1, Y2, and Y3. Because a causal path fixed at zero is equivalent to omitting the path, to reduce clutter I omit the path from 'Slope' to Y0 in Figure 16.17. I elaborate on these rules of thumb later after I provide more background. Table 16.12 provides the Mplus syntax to apply this model to the data.

Table 16.12. Mplus Syntax for Classic Latent Growth Model

```
1.  TITLE: Classic LGM  ;
2.  DATA: FILE = LGAclassicM.dat;
3.  VARIABLE:
4.    NAMES = y0 y1 y2 y3;
5.  ANALYSIS: ESTIMATOR=MLR ;
6.  MODEL:
7.    ! Define latent growth model: i=latent intercept, s=latent slope
8.    ! Equidistant time points are (0, 1, 2, 3)
9.    i s | y0@0 y1@1 y2@2 y3@3;
10.   ! Estimate means of latent intercept and latent slope
11.   [i];      ! Mean intercept
12.   [s];      ! Mean slope
13.   !Estimate variances of latent intercept and slope
14.   i;       ! Variance of intercept
15.   s;       ! Variance of slope
16.   !Estimate the correlation between the latent intercept and slope
17.   i WITH s;
18.   ! Estimate the disturbance variances (error)
19.   y0-y3;
20. OUTPUT: Samp StdYX MOD(4) Residual Tech4  ;
21. PLOT: TYPE=PLOT3 ;
22.       SERIES = Y0(0) Y1(1) Y2(2) Y3(3) ;
```

Line 9 is a programming shortcut built into Mplus. To the left of the | are the names you want to assign to the latent intercept and the latent slope, in that order. Each name can be up to 8 characters in length. I label them `i` and `s`, respectively. Mplus internally assigns all of the loadings for the intercept factor to 1.0 and assigns the slope loading value listed next to each variable followed by a @ to the right of the |. In this case, 0 is assigned to the slope loading for Y0, 1 is assigned to the slope loading for Y1, 2 is assigned to the slope loading for Y2, and 3 is assigned to the slope loading for Y3. The rest of the code should be self-explanatory when you also read the comment lines associated with them. The one exception is the `PLOT` command, which asks Mplus to generate plots for the model. For growth curves, you always set the type of plot to `TYPE3`. The `SERIES` subcommand specifies the sequence of the repeated measure variables to use on the x-axis. In this case, Y0 will be first, Y1 will be second, Y2 will be third, and Y3 will be last. You access the plots after executing the syntax in Table 16.12 and then clicking the Plot menu item at the top of the Mplus interface. The formatting of the plots is a bit crude so I often rely on the plot programs on my website.

When I executed the syntax, the output provided the traditional indices of model fit. The chi square test of perfect model fit in the population was statistically non-significant (chi square = 2.66 with 5 degrees of freedom, $p < 1.00$), which is consistent with the

proposition that the data are reasonably in accord with the model. The RMSEA was less than 0.001. The upper limit of the 90% confidence interval for it is 0.009. The p value for close fit was not statistically significant (p < 1.00). The CFI is 1.00 and the standardized RMR was 0.003. For localized fit, there were no theoretically meaningful modification indices greater than 4 and no meaningful residuals between the predicted and observed covariances on a cell by cell basis. Overall, model fit seems reasonable.

Here are the results for the core model statistics that are of substantive interest:

```
                                                    Two-Tailed
                   Estimate       S.E.  Est./S.E.    P-Value

 Means
    I                 0.522      0.009     58.769      0.000
    S                 0.199      0.004     45.592      0.000

Variances
    I                 0.481      0.013     35.926      0.000
    S                 0.102      0.003     30.264      0.000
```

The mean of the intercept is the estimated mean Y at time 0. It was 0.52 ±0.02 on the -5 to +5 Y metric.[7] The average slope across individuals for Y was 0.20 ±0.008 (CR = 45.59, p < 0.05). This represents the "typical" trajectory for individuals as represented by a linear function. For each unit that lag increased, the mean of Y increased by 0.20. The predicted mean at the first assessment (time 0) was 0.52, at the first lag (time 1) it was 0.52 + 0.20 = 0.72, at the second lag (time 2) it was 0.72 + 0.20 = 0.92, at the third lag (time 3) it was 0.92 + 0.20 = 1.12. Thus, the typical growth was for Y to increase linearly by 0.20 every 6 months.

The variance reported below the `Means` section provides a sense of how much variability there is across individuals about the "typical" slope. The variance for the slopes was 0.102 and the standard deviation of the slopes is the square root of this or 0.32. This represents quite a bit of variability, suggesting that we might want to explore in a substantive sense (perhaps in another study) what causes such variability. Parenthetically, the standard error and significance tests for variances and significance tests (reported above) can be biased under the standard statistical theory underlying growth curve modeling. Bootstrapping is therefore recommended for evaluating this facet of model. You can readily apply bootstrapping in Mplus as discussed earlier by changing the command on Line 5 of Table 16.12 to

[7] Technically, the latent mean intercept often will not equal the exact value of the observed mean at baseline because the intercept is an estimated latent factor representing the *modeled* population average at time zero, while the observed baseline is a raw sample mean influenced by measurement error and missing data.

```
ANALYSIS: ESTIMATOR=ML ; BOOTSTRAP=2000 ;
```

I usually report the robust maximum likelihood solution but also inform readers that I conducted sensitivity analyses using bootstrapping.

The above analysis defined the linear slope using the metric of the number of lags, namely by fixing the loadings in Figure 16.17 to 0, 1, 2, 3. The lag between measures was six months, so the resulting mean slope is interpreted as how much the mean of Y is predicted to change every 6 months, namely 0.20 ±0.008. Suppose I wanted to express the slope on a per month basis rather than a six-month basis. I can do so by setting the loadings to 0, 6, 12, 18 instead of 0, 1, 2, 3. Note that a linear function is preserved but I have rescaled the lags to months by multiplying them by 6. When I execute the modified syntax, the global fit indices are unchanged and this also is true for the critical ratios and p values for all the coefficients of interest. However, the mean slope is now reported as 0.033 ±0.002 such that the monthly change in the mean Y is predicted to be 0.033. Note that if I multiply this value by 6. I obtain 0.20, which comports with the original analysis. I can use this "trick" to obtain estimates using any time unit I desire. To obtain per day estimates that assume 30 days in a month, I would use the coefficients 0, 30, 60, and 90.

Manipulating the loading coefficients also can deal with unequal lags while maintaining a linear function. For example, suppose the time between lag 2 and 3 was 12 months instead of 6 months but all of the other lags were 6 months. Instead of using loadings of 0, 1, 2, and 3 I would use loadings of 0, 1, 2, and 4, with the time between the last two lags of 2 and 3 representing two six-month lags rather than only one lag. Or, I could use the loadings 0, 6, 12, and 24. As I elaborate in the document *Additional Applications of Latent Growth Curve Modeling* on my webpage, manipulating the values of the loadings leads to flexibility in the models you can test.

The Latent Growth Curve Model in RETs

With the above as background, I can introduce more complex latent growth curve (LC) models that can be applied to RETs. There are different approaches to implementing such applications with the primary candidates being (a) a single group approach that incorporates a dummy variable for the treatment condition, and (b) a multiple group approach that applies multi-group SEM treating the intervention and control groups as separate populations. Muthén, and Curran (1997) advocate the use of the multigroup strategy but acknowledge that the dummy variable approach also is appropriate in many contexts. I illustrate the single group approach here but I also present a detailed example using the multi-group strategy in the document on my webpage *Additional Applications*

of Latent Growth Curve Modeling. For an application of the dummy variable approach using binary outcomes in an intervention context, see Muthén & Muthén (2008).

For pedagogical reasons, I restrict the example to the case of a single mediator but it is easily extended to multiple mediators. The RET seeks to compare two treatments, A and B, so instead of intervention and control groups, the dummy variable TREAT is scored 0 = Treatment B and 1 = Treatment A. The researcher believes that Treatment A will prove to be superior to Treatment B in influencing the outcome, Y, because it will more strongly impact the presumed mediator of the treatment effect, M, than Treatment B. Higher scores on M and higher scores on Y are desirable. The intervention sought to increase them both. The mediator, M, and the outcome, Y, were each measured at five time points. The baseline measures, M0 and Y0, were obtained prior to the start of treatment with M1 and Y1 representing scores at the immediate posttest. The five assessments of the mediators are referred to as M0 M1, M2, M3, and M4 and the Ys are referred to as Y0, Y1, Y2, Y3, and Y4. After the measurement of M1 and Y1, the three follow-ups were each lagged by 6 months. The metrics of all the variables were such that their standard deviations approximated values of 1.0. The mediator was parameterized as a growth curve model as was the outcome, so the full model contains two growth curves, one for the mediator and one for the outcome. The hypothesis is that the growth curve parameters on Y are linked to the growth curve parameters on M, a design known as a **parallel process latent growth curve design** (Cheong, MacKinnon & Khoo, 2003). The relevant influence diagram is in Figure 16.18. Correlated disturbances between d5 and d6 and between d7 and d8 are omitted from the diagram to reduce clutter but they were included in the Mplus syntax for purposes of testing the model. These correlations reflect the fact that the respective latent intercepts and slopes for M and Y respectively may be correlated over and above the determinants of them explicitly shown in the model.

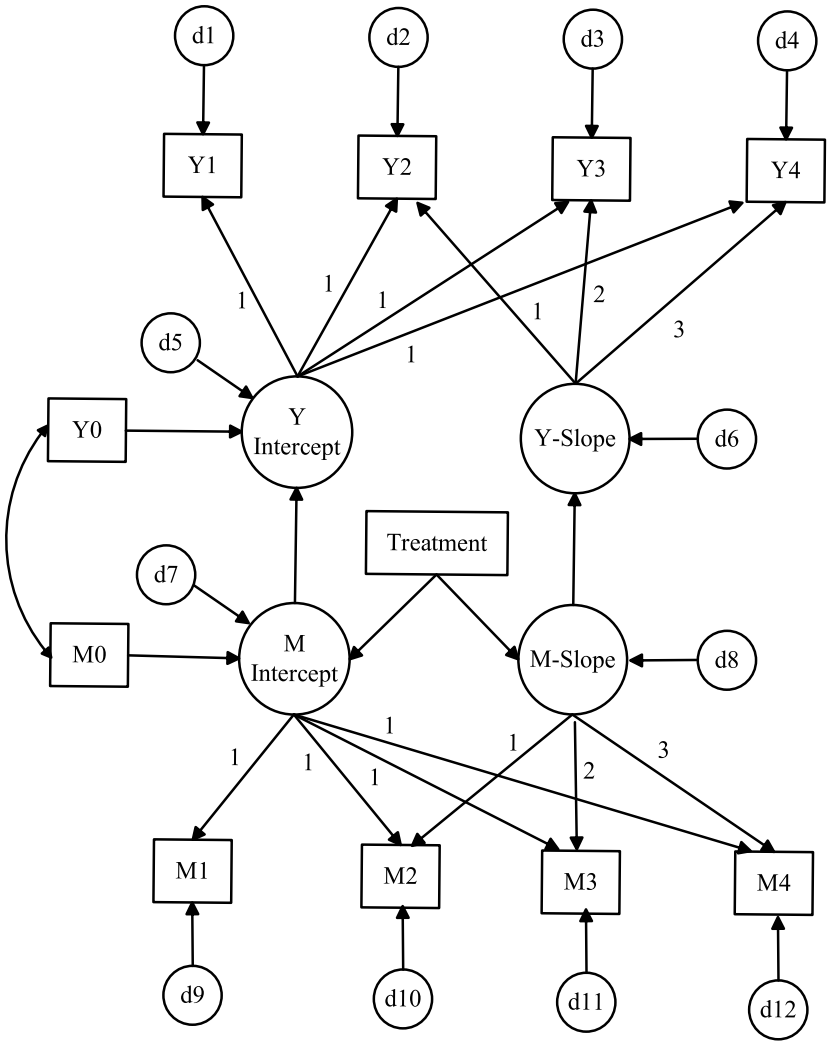
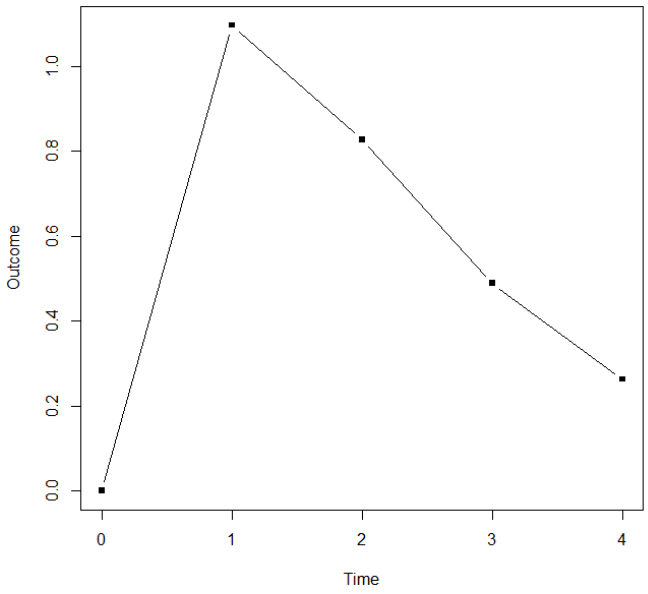


FIGURE 16.18 RET growth curve model

There are several unique features of the model. First, note that I have not included the baseline measures, M0 and Y0, as part of the growth curves *per se*. The curves instead begin with the immediate posttest and extend through the last follow-up, namely M1 through M4 and Y1 through Y4. Many applications of linear growth curve modeling to RCTs include the baseline as part of the curve, but doing so almost certainly necessitates a non-linear rather than a linear growth function to model. This is because the mean difference in Y between the baseline and the immediate posttest will almost certainly be upward and positive (assuming higher scores on Y are the goal of the intervention) but the curve thereafter from Y1 to Y4 will be downward and negative as

the initial change from baseline to the immediate posttest decays over time. Here is the plot of the observed Y means across the full time range for the current example:



The curve is clearly non-linear. Forcing a linear function onto the growth curve using the *a priori* defined fixed loadings for the slope factor represents a non-trivial misspecification. When the analysis is executed in Mplus, careful examination of model diagnostics (e.g., global and localized fit indices) should reveal the misspecification. Note that if I remove the baseline Y from the above plot, we are left with a downward slope that closely approximates linearity between Y1 and Y4. The mean slope parameter in the "growth" curve analysis, as shown later, is then analogous to a decay curve.

One way of thinking about this approach invokes classic growth curve logic for mapping how a variable of interest changes across time after exposure to a major life event, such as a death in the family, getting married, having a child, or retiring from work. One defines the onset of the event and then maps how people react immediately after and then with the passage of time up to some endpoint. In an RCT or RET, one can conceptualize the starting point or event onset as exposure (or non-exposure) to the intervention. We document where people stand on the outcome just after exposure or non-exposure to the intervention and then map how the outcome decays or grows over the course of the ensuing days, weeks, months, or years for different groups of people The different groups can be those exposed to the intervention versus those in the control condition or, in the present case, those exposed to Treatment A versus those exposed to Treatment B. This logic justifies the model frame I propose here. Note in this framework

that the mean intercept for Y in the growth curve model reflects the estimated mean Y value at the immediate posttest, i.e., the start of the developmental process following event exposure. The value of this mean might be expected to differ for the two treatment groups and the question then turns to how this difference changes or evolves thereafter across time per the trajectories over time.

A second noteworthy aspect of the model is my use of the baseline measures of Y and M as covariates for the latent mean intercepts of Y and M, respectively. This is consistent with the standard RCT practice of using baseline measures as covariates to increase statistical power for comparisons at the immediate posttest and to adjust for incidental condition imbalance associated with random assignment. The strategy of using the baseline measure of Y as a covariate instead of including it as a formal part of the growth curve in randomized trials is empirically supported by Chou et al. (2010).

The Mplus syntax for the analysis appears in Table 16.13.

Table 16.13. Mplus Syntax for RET Example

```
TITLE: Parallel Process RET LGM FOR TWO TREATMENT GROUPS  ;
DATA: FILE = LGADATAM.dat;
VARIABLE:
DEFINE:
 CENTER m0 y0 (GRANDMEAN) ;
NAMES = y1 y2 y3 y4 m1 m2 m3 m4 y0 m0 treat ;
ANALYSIS: ESTIMATOR=MLR ;
MODEL:
 !Define latent growth model for y
 yi ys | y1@0 y2@1 y3@2 y4@3 ;
 !Estimate latent intercepts and assign them labels
 [yi] (ya1);      !Intercept for intercept factor
 [ys] (ya2);      !Intercept for slope factor
 !Estimate variances of intercept and slope factors
 yi;       ! Variance of intercept factor
 ys;       ! Variance of slope factor
 !Estimate covariance between intercept and slope factors
 yi WITH ys;
 ! Estimate residual variances (error or disturbances)
 y1-y4;
 !Define latent growth model for mediator
 mi ms | m1@0 m2@1 m3@2 m4@3 ;
 !Estimate intercepts and assign them labels
 [mi] (ma1) ;      !Intercept for intercept factor
 [ms] (ma2) ;      !Intercept for slope factor
 !Estimate variances of intercept and slope factors
 mi;       ! Variance of intercept factor
 ms;       ! Variance of slope factor
 !Estimate covariance between intercept and slope factors
```

```
31.  mi WITH ms;
32.  !Estimate residual variances (error)
33.  m1-m4;
34.  !Define the regressions and assign labels to coefficients
35.  mi ON treat m0 (p1 p2)  ;
36.  ms ON treat (p3);
37.  yi ON mi y0 (p4 p5);
38.  ys ON ms (p6);
39. MODEL INDIRECT:
40.  yi IND treat ;
41.  mi IND treat ;
42.  ys IND treat ;
43.  ms IND treat ;
44. !OUTPUT: Samp StdYX MOD(4) Residual Tech4  ;
```

All of the syntax should be self-explanatory at this point except for my addition of some `MODEL INDIRECT` commands. Lines 39 to 43 request that Mplus calculate the effects of the treatment condition on the two latent factors for the mediator and the two latent factors for the outcome. As you will see, these will be useful to address the three standard RET questions (is there an overall treatment effect on the outcome, is there a treatment effect on the mediators, and is there an effect of the mediators on the outcome).

When I executed the syntax in Table 16.13, the model fit indices all suggested good model fit. The chi square test of perfect model fit in the population was statistically non-significant (chi square = 46.77 with 44 degrees of freedom, $p < 0.999$), which is consistent with the data being in accord with the model. The RMSEA was 0.008. The upper limit of the 90% confidence interval for it is 0.023. The p value for close fit was not statistically significant ($p < 1.00$). The CFI is 1.00 and the standardized RMR was 0.019. For localized fit, there were no theoretically meaningful modification indices greater than 4 and no meaningful residuals between the predicted and observed covariances on a cell-by-cell basis.

After concluding in favor of a reasonable model fit, I re-ran the syntax but added some `MODEL CONSTRAINT` commands. The reason I did not include these in the initial analysis is that Mplus does not allow for the reporting of modification indices when the `MODEL CONSTRAINT` command is used. I needed access to the modification indices to properly evaluate model fit, so I delayed adding the commands until after I concluded the model fit was reasonable. For the new program, I replaced line 44 with the line:

```
OUTPUT: Samp StdYX Residual Tech4  ;
```

Here are the commands I added just before the new Line 44:

```
43a1.  MODEL CONSTRAINT:
```

```
43a2.    NEW (msta mstb msdiff m1ta m2ta m3ta m4ta m1tb
43a3.    m2tb m3tb m4tb mdiff1 mdiff2 mdiff3 mdiff4) ;
43a4.    msta = ma2 + p3 ; !Treatment A mean of mediator slope factor
43a5.    mstb = ma2 ;      !Treatment B mean of mediator slope factor
43a6.    msdiff = msta-mstb ;
43a7.    m1ta = (ma1+p1) ;       !Treatment A mediator value at time 1
43a8.    m2ta = m1ta + msta*1  ; !Treatment A mediator value at time 2
43a9.    m3ta = m1ta + msta*2  ; !Treatment A mediator value at time 3
43a10.   m4ta = m1ta + msta*3  ; !Treatment A mediator value at time 4
43a11.   m1tb = ma1   ;          !Treatment B mediator value at time 1
43a12.   m2tb = m1tb + mstb*1 ;  !Treatment B mediator value at time 2
43a13.   m3tb = m1tb + mstb*2 ;  !Treatment B mediator value at time 3
43a14.   m4tb = m1tb + mstb*3 ;  !Treatment B mediator value at time 4
43a15.   mdiff1 = m1ta - m1tb ;
43a16.   mdiff2 = m2ta - m2tb ;
43a17.   mdiff3 = m3ta - m3tb ;
43a18.   mdiff4 = m4ta - m4tb ;
43b1.    NEW (ysta ystb ysdiff y1ta y2ta y3ta y4ta y1tb
43b2.    y2tb y3tb y4tb ydiff1 ydiff2 ydiff3 ydiff4) ;
43b3.    ysta = ya2 + (ma2 + p3)*p6 ; !Treatment A mean of Y slope factor
43b4.    ystb = ya2 + ma2*p6 ;        !Treatment B mean of Y slope factor
43b5.    ysdiff = ysta-ystb ;
43b6.    y1ta = ya1 + (ma1 + p1)*p4;  !Treatment A Y value at time 1
43b7.    y2ta = y1ta + ysta*1;        !Treatment A Y value at time 2
43b8.    y3ta = y1ta + ysta*2;        !Treatment A Y value at time 3
43b9.    y4ta = y1ta + ysta*3;        !Treatment A Y value at time 4
43b10.   y1tb = ya1 + ma1*p4;         !Treatment B Y value at time 1
43b11.   y2tb = y1tb + ystb*1;        !Treatment B Y value at time 2
43b12.   y3tb = y1tb + ystb*2;        !Treatment B Y value at time 3
43b13.   y4tb = y1tb + ystb*3;        !Treatment B Y value at time 4
43b14.   ydiff1 = y1ta - y1tb ;
43b15.   ydiff2 = y2ta - y2tb ;
43b16.   ydiff3 = y3ta - y3tb ;
43b17.   ydiff4 = y4ta - y4tb ;
```

Lines 43a4 to 43a6 use the labels from the syntax in Table 16.13 to calculate the mean slope for Treatment A, the mean slope for Treatment B, and the difference between them. Lines 43a7 to 43a10 calculate the predicted mean of the mediator for Treatment A at lags 1, 2, 3, and 4, respectively. Lines 43a11 to 43a14 calculate the predicted mean of the mediator for Treatment B at lags 1, 2, 3,and 4, respectively. Lines 43a15 to 43a18 calculate the difference between these means for Treatment A minus Treatment B. Lines 43b1 to 43b18 calculate comparable parameters but for the outcome not the mediator.

I do not need to calculate any of the above parameters beyond the information that is reported on the traditional output unless I want to be more thorough in my presentation of results. Appendix A outlines the logic for each of the `MODEL CONSTRAINT` expressions.

I now consider the core questions of the RET but adapted to the comparison of the two treatments instead of the more traditional treatment versus control conditions. I comment on more traditional designs in a later section.

Relative Effects of the Treatments on the Outcome

To test the relative effects of Treatment A versus Treatment B on the outcome, I use the output generated by Lines 40 (`YI IND TREAT`) and 42 (`YS IND TREAT`) from Table 16.13. The former focuses on the treatment mean difference at the immediate posttest and the latter focuses on the treatment differences in decay in Y that occurs across the follow-ups. Here is the output for Line 40 that focuses on the difference between Treatment A minus Treatment B at the immediate posttest:

```
TOTAL, TOTAL INDIRECT, SPECIFIC INDIRECT, AND DIRECT EFFECTS

                                                    Two-Tailed
                    Estimate       S.E.  Est./S.E.    P-Value

Effects from TREAT to YI

  Total                0.035      0.019      1.871      0.061
  Total indirect       0.035      0.019      1.871      0.061

  Specific indirect 1
    YI
    MI
    TREAT              0.035      0.019      1.871      0.061
```

The entry in the first column underneath `Specific indirect 1` identifies the causal chain(s) through which the treatment condition reaches the Y latent intercept. It is `TREAT→MI→YI`. The row of primary interest is the one called `Total` underneath the `Effects from TREAT to YI` label. The estimate in this row is the predicted difference between Y at the immediate posttest for Treatment A minus Y at the immediate posttest for Treatment B. In this case, the covariate adjusted difference is 0.035 ±0.038, which is statistically nonsignificant (CR = 1.87, $p < 0.07$). Suppose that prior to the study, the research team set a standard for a meaningful mean difference at the immediate posttest as a population absolute difference of 0.20 or greater. The criterion for defining a truly trivial difference, i.e., one that is functionally zero via the latitude of no effect (see Chapter 10) was a population absolute value less than 0.10. Stated another way, the interval for the latitude of no effect was -0.10 to 0.10. The 95% confidence interval for the treatment difference was -0.03 to 0.073 which is fully contained within the latitude of

no effect. I can therefore conclude that, functionally, there is no difference in the population mean Y at the immediate posttest for the two treatments.

If desired, I also can report the predicted means of Y at the immediate posttest for Treatment A and for Treatment B to embellish the above results. These predicted values are labeled `Y1TA` and `Y1TB` in the `MODEL CONSTRAINT` command. Here is the output:

```
New/Additional Parameters
                                                    Two-Tailed
                    Estimate       S.E.  Est./S.E.    P-Value

    Y1TA               1.113      0.033     34.127      0.000
    Y1TB               1.079      0.032     33.761      0.000
```

The estimated Y means for Treatment A and Treatment B at the immediate posttest were 1.113 ± 0.06 and 1.079 ± 0.06, respectively.

Next, I examined a second facet of the relative effects of the two treatments on the outcome, namely whether they had meaningfully different decay curves. To do so, I examined the comparable section to the prior one but now for the Y latent slope, `YS`. The relevant output is associated with `YS IND TREAT`. Here is the core output:

```
Effects from TREAT to YS

  Total               0.077      0.015      5.135      0.000
  Total indirect      0.077      0.015      5.135      0.000

  Specific indirect 1
    YS
    MS
    TREAT             0.077      0.015      5.135      0.000
```

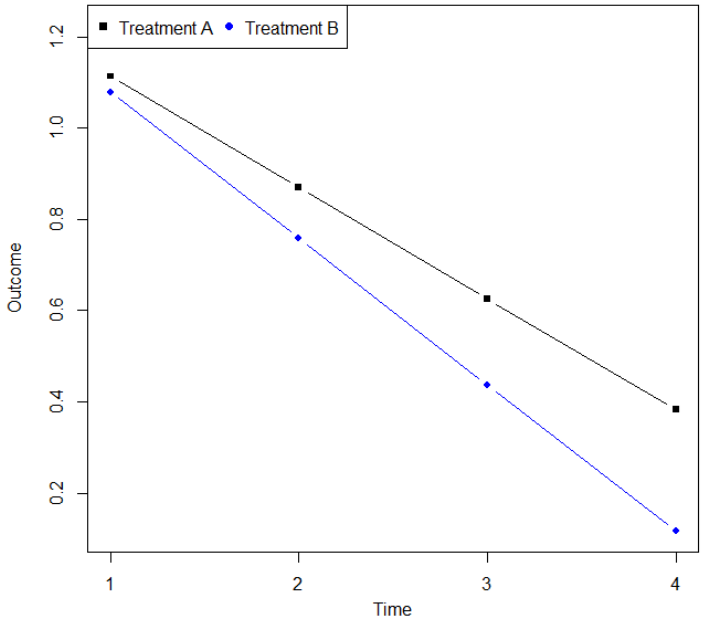
The row of primary interest is the one called `Total` underneath the `Effects from TREAT to YS` label. The estimate in this row is the predicted mean slope difference between Treatment A minus Treatment B for linear decay from the immediate posttest for Y to the value of Y eighteen months later. The predicted mean difference in the decay curve for Treatment A minus Treatment B was 0.077 ±0.030, which was statistically significant (CR = 5.14, $p < 0.05$). Suppose that prior to the study, the research team set a standard for a meaningful slope difference as an absolute difference of 0.04 or greater. The confidence interval for the slope difference was 0.05 to 0.18. Because the lower limit of the interval exceeds the standard, I conclude the difference is meaningful.

I might also want to report the separate values of the decay curve/slope for Treatment A and for Treatment B. These values are labeled YSTA and YSTB in the output section from the MODEL CONSTRAINT commands. Here is the edited output:

```
New/Additional Parameters
                                                    Two-Tailed
                    Estimate       S.E.  Est./S.E.    P-Value

    YSTA              -0.244      0.017    -14.425      0.000
    YSTB              -0.321      0.017    -19.290      0.000
```

The decay is weaker in Treatment A than in Treatment B.

I can plot the two curves using the information in the output section from the MODEL CONSTRAINT commands for the predicted means for Y1 though Y4 for Treatment A (labeled Y1TA, Y2TA, Y3TA, and Y4TA) and for Treatment B (labeled Y1TB, Y2TB, Y3TB, and Y4TB). Using the program on my website called *Temporal line plot*, here is the plot:



The decay/slope for Treatment B is steeper than that for Treatment A, which is to be expected given the results reported earlier. The separation between lines at a given time point reflects the predicted mean difference between Treatment A and Treatment B at that time point. It can be seen that the mean difference becomes more marked as time passes. This is because the mean Y is decaying more slowly for Treatment A than for Treatment B.

In sum, the two treatments have comparable effects on Y at the immediate posttest. However, Treatment A is favored over treatment B, everything else being equal, because it has a more favorable decay curve. Parenthetically, the mean Y at baseline collapsing across the two treatment conditions was 0.00 ±0.06 with an SD of 1.03. The immediate posttest means for both treatments were substantially larger than this baseline value (both were near 1.00), suggesting non-trivial effects of both treatments. However, such interpretations must be made with caution because the experimental design lacks a control condition that controls for such time-varying factors like history effects, maturation, and the like (see Chapter 4). The study would be stronger if it included a control condition so we could then characterize effect sizes of each treatment relative to a control group rather than only to each other. As a study that compares the equivalence or superiority of one treatment over another, the design is fine. However, formal statements about the effect size of each treatment per se are probably better made with the addition of a control group.

Relative Effects of the Treatments on the Mediator

The analysis of the relative effects of the two treatments on the mediator follows the same structure as that for the total relative effect on Y. I use the output generated by Lines 41 (`MI IND TREAT`) and 43 (`MS IND TREAT`) from Table 16.13. Here is the output for Line 41 that focuses on the mediator difference between Treatment A minus Treatment B at the immediate posttest:

```
TOTAL, TOTAL INDIRECT, SPECIFIC INDIRECT, AND DIRECT EFFECTS

                                                    Two-Tailed
                    Estimate       S.E.  Est./S.E.    P-Value

Effects from TREAT to MI

  Total                0.115      0.058      1.973      0.048
```

The row of interest is the `Total` row, which is the predicted difference between M at the immediate posttest for Treatment A minus Treatment B. In this case, the covariate adjusted difference is 0.115 ±0.114, which is statistically significant (CR = 1.97, $p < 0.05$). Suppose that prior to the study, the research team set a standard for a meaningful mean difference in M at the immediate posttest as an absolute difference of 0.20 or greater. The criterion for defining a truly trivial difference, i.e., one that is functionally zero via the latitude of no effect was an absolute value less than 0.10 or a latitude of no effect of -0.10 to 0.10. The 95% CI for the treatment difference was -0.001 to 0.229. This

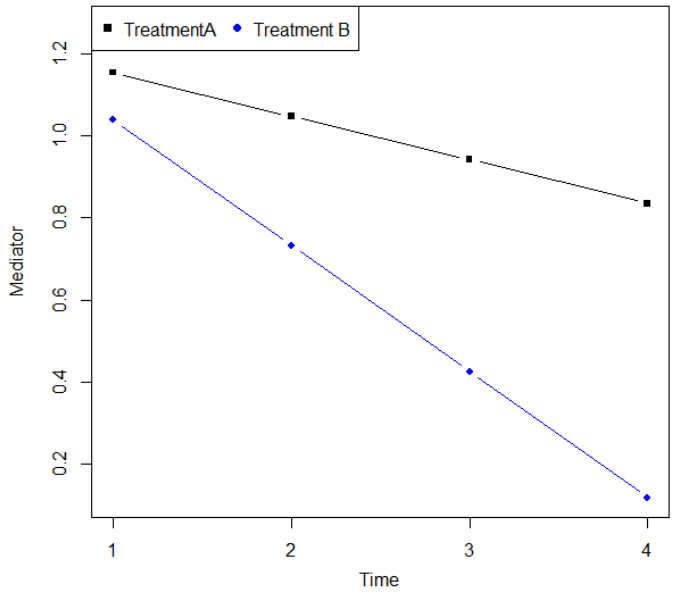
overlaps the meaningfulness latitude as well as the latitude of no effect. Although the result favors Treatment A over Treatment B in terms of statistical significance, we cannot draw firm conclusions in terms of difference meaningfulness.

If desired, I also can report the values of the predicted mediator means at the immediate posttest for Treatment A and for Treatment B. These values are labeled M1TA and M1TB in the output section from the MODEL CONSTRAINT commands. The estimated means for Treatment A and Treatment B at the immediate posttest were 1.15 ± 0.09 and 1.04 ± 0.08, respectively.

Next, I address treatment differences in the decay curves by examining the output from the MS IND TREAT command on Line 43. Here is the relevant output:

```
                                                    Two-Tailed
                   Estimate       S.E.  Est./S.E.    P-Value

Effects from TREAT to MS

Total                 0.201      0.028      7.208      0.000
```

For the mediator, the predicted mean difference in the decay curve for Treatment A minus Treatment B was 0.201 ±0.056, which was statistically significant (CR = 7.21, $p < 0.05$). Suppose that prior to the study, the research team set a standard for a meaningful slope difference as an absolute difference of 0.04 or greater. The confidence interval for the slope difference was 0.15 to 0.26. Because the lower limit of the interval exceeds the meaningfulness standard, I can conclude that the slope difference between treatments for the mediator is meaningful. The separate values of the decay curves for Treatment A and for Treatment B are located in MSTA and MSTB in the output section from the MODEL CONSTRAINT commands. For Treatment A, the slope was -0.106 ± 0.04 and for Treatment B it was -0.307 ± 0.04. The decay was more rapid for Treatment B than for Treatment A, meaningfully so. If desired, I can plot the two curves, as I did before with Y (the relevant information for the plot is in M1TA, M2TA, M3TA, M4TA, M1TB, M2TB, M3TB, and M4TB) in the MODEL CONSTRAINT command. Here is the plot using the program *Temporal line plot* from my website:



The slope in Treatment B is steeper than the slope in Treatment A, with mean differences on the mediator becoming more pronounced as time passes.

In sum, I could not confidently conclude that one of treatments had a meaningfully superior effect than the other treatment on the mediator at the immediate posttest. However, Treatment A is favored over treatment B, everything else being equal, because it has a more favorable decay curve with respect to the mediator.

Effect of the Mediator on the Outcome

The analysis of the estimated effect of the mediator on the outcome does not distinguish the two treatments because the guiding model in Figure 16.12 assumes that the effect is the same in both treatment conditions. I describe methods for testing this assumption in the document *Additional Applications of Latent Growth Curve Modeling* on my website.

The relevant coefficients for evaluating the mediator effect on the outcome are the coefficient from `MI` to `YI` and from `MS` to `YS`. Here are the results from the output:

```
MODEL RESULTS

                                                    Two-Tailed
                  Estimate       S.E.  Est./S.E.    P-Value

YI       ON
    MI             0.302      0.045      6.672      0.000
    Y0             0.093      0.028      3.296      0.001
 YS       ON
    MS             0.385      0.055      7.016      0.000
```

The coefficient for MI→YI is the path coefficient for the outcome at the immediate posttest regressed onto the mediator at the immediate posttest controlling for the baseline outcome. The coefficient was 0.30 ± 0.09 (CR = 6.67, $p < 0.05$). For every one unit that the mediator increased across individuals, the outcome was predicted to increase by 0.30 units. Suppose the research team decided prior to the study that a coefficient of 0.20 or greater would be deemed meaningful. The 95% confidence interval for the coefficient was 0.21 to 0.39. Because the lower limit of this interval is greater than meaningfulness standard, we can declare the result meaningful.

The coefficient for MS→YS is the path coefficient for the decay slope of the outcome regressed onto the decay slope for the mediator. A positive coefficient implies that people with increasingly more positive slopes on the mediator tend to have increasingly more positive slopes on the outcome, i.e., that there is an association between the two. The path coefficient was 0.385 ± 0.11 (CR = 7.02, $p < 0.05$). For every one unit that the mediator slope increased across individuals, the outcome slope was predicted to increase by 0.385 units. Suppose the research team decided prior to the study that a coefficient of 0.20 or greater would be deemed meaningful. The 95% confidence interval for the coefficient was 0.27 to 0.49. Because the lower limit of this interval is greater than the meaningfulness standard, we can declare the result meaningful.

Note that the only covariate I included in these analyses was the baseline Y0 for YI. I did so to keep the example simple. In practice you likely would have more covariates. The covariates can be time-invariant or time-varying. In the current case, I treated the baseline Y as a time-invariant covariate. Time varying covariates are modeled as growth curves and introduced into the model accordingly. I discuss alternative approaches to incorporating time-varying covariates in the document *Additional Applications of Latent Growth Curve Modeling* on my website.

In my analyses, I did not include a direct effect of the treatment condition on YI or YS over and above the effect of the treatment condition on MI and MS. This presumes that such a direct effect is functionally zero. If desired, a researcher can include such effects and many methodologists routinely do so. In the current case, the research team saw no conceptual reason to include them. When I did so, both the estimated direct effect of the treatment on YI (coefficient = 0.03 ± 0.12, CR = 0.48, *ns*) and on YS (coefficient = -0.004 ± 0.06, CR = 0.13, *ns*) were trivial. The chi square test of fit for the model that included them was 46.53, df = 42, $p < 0.30$). A nested chi square difference test with the model that excluded the two effects yielded a statistically nonsignificant chi square difference of 1.24, df = 2, *ns*). For a discussion of the inclusion/exclusion of direct effects independent of mediators more generally, see my discussion of the Baron-Kenny mediation method in Chapter 9.

Omnibus Mediation Analysis

As in previous chapters, omnibus mediation tests are of lower priority for me. I instead prefer to focus on the individual links in mediational chains, evaluating their effects sizes. If I want an overall omnibus test of statistical significance, I tend to rely on the joint significance test, per earlier chapters.

The current RET compared two treatments without a control group, so statements of absolute treatment effects are on weak grounds. In LGC models, evaluations of mediation takes two forms. The first focuses on treatment differences in the latent intercept for the mediator and whether these differences, in turn, translate into treatment differences in the outcome latent intercept. Given there were no meaningful differences between the two treatment conditions on the mediator, this link in the mediational chain for the latent intercept was "broken." Indeed, there were no meaningful effects of the treatment condition on the Y latent intercept so there is not much to explain in the first place.

The second mediation form focuses on the latent slopes of the mediator and the outcome. This essentially extends mediation analysis to one on treatment decay. The RET found meaningful and statistically significant treatment differences in the latent slopes for the mediator and meaningful and statistically significant estimates of the causal coefficients linking the mediator slope to the outcome slope. This is consistent with omnibus mediation in this second mediation analysis facet.

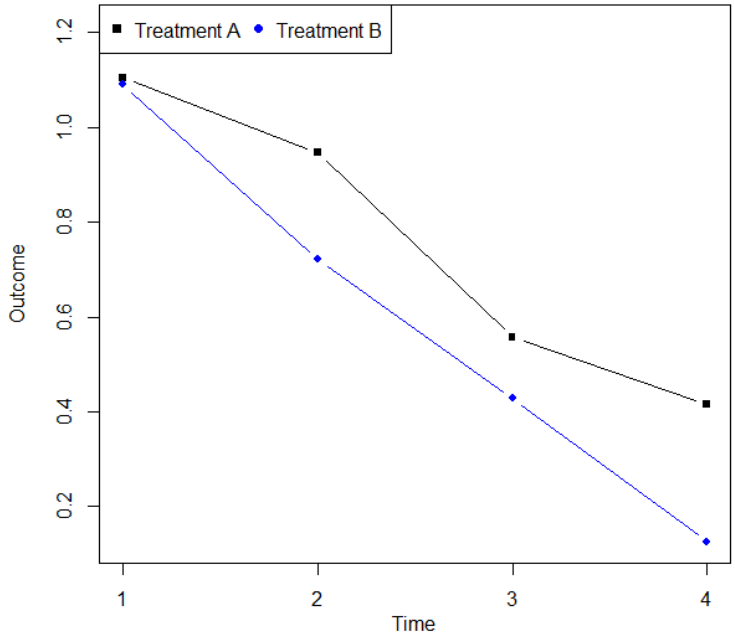
Concluding Comments for the Example

The growth curve analyses in the current example rest on several core assumptions of the modeling technique. The first is that the growth function, in the current case, is linear in form. A second assumption is that the observed data are generated by latent factors representing the initial status (intercept) and the rate of change (slope). It often is assumed these latent factors are multivariate normally distributed. The use of robust estimators (such as MLR) reduces the need for normality assumptions. Third, the model assumes the individual cases in the analysis are independent. The multiple measurements within individuals do not make this assumption. Indeed, they typically are not independent. LGC modeling does not assume homogeneity of disturbances across time unlike traditional repeated measures ANOVA.

Like any RET, the results of LGC modeling can be impacted by measurement error and omitted variable bias. Your conclusions should be tempered accordingly. Also, the tested model in this example assumed contemporaneous effects of the mediator on the outcome rather than lagged effects. Lagged effects can be dealt with. For example, one can predict an outcome at time point t using the mediator at time point $t-1$ by parameterizing the mediator as such. In the document *Additional Applications of Latent*

Growth Curve Modeling on my website, I address ways of incorporating lagged effects, measurement error, and omitted variable bias as well as sensitivity tests relative to them.

I want to stress the importance of correctly specifying the growth function in one's model. The current example assumed a linear one. If the growth function is misspecified, then the fit of the SEM model should be poor and reflected in fit diagnostics. This is an advantage of using SEM for growth curve analysis. I like to augment the diagnostics with other checks when possible. For example if I plot the observed Y means (not the predicted means) across the different time points for the two treatment groups, do I observe reasonable approximations to linearity? Using the *Temporal line plot* program on my website, here is the plot for Y from the immediate posttest to the 18 month follow-up:



The trends seem close enough to linearity. This also was true when I plotted the mediator observed means. Granted, there are some deviations from linearity in the above plot but the deviations likely are due to random noise. The plot differs from those I presented earlier because the prior plots are model-based, i.e., they assume the fitted model faithfully represents the true data generating process. This is why we want to ensure that our hypothesized model has good model fit and is conceptually viable.

Concluding Comments/Extensions of the Latent Growth Curve Model

The method of analysis I outlined focuses on how trajectories for different variables are

related to one another, in this case how the trajectory for a mediator is related to the trajectory for an outcome. This is a common form of mediational analysis in LGC models. The model can be extended in multiple ways including to the case of multiple mediators, non-linear growth functions, the analysis of traditional latent variables to adjust for measurement error, different approaches to the inclusion of covariates, and the analysis of binary and ordinal outcomes, to name a few. I discuss these applications in the document *Additional Applications of Latent Growth Curve Modeling* on my website as well as modeling alternatives that retain the spirit of LGC frameworks. I also illustrate an analysis of the current RET example but with a control group added as well as traditional two group design with an intervention and control group.

In general, for RETs I prefer the SEM-based fixed effects analytic strategies I described in prior sections of this chapter to LGC modeling. The former generally are more powerful, provide more information, and are easier to implement. If you are truly interested in trajectories *per se*, especially condition differences in decay functions, then the LGC framework is a reasonable choice. Keep in mind, however, that growth curve modeling yields coefficients that describe variability *between* individuals but for within-person coefficients that reflect change over time. At heart, LGC modeling is an across-person form of analysis.

The parallel process mediational strategy I introduced tends to be sample size demanding (Cheong, 2011). Generally speaking, power and accuracy of the method improve with increases in sample size, effect size of the tested effects, and the number of measurement occasions. Depending on these factors, sample sizes as low as 150 but as large as 1,000 might be needed. I encourage you to use the simulation strategies for power analysis outlined in Chapter 28 to explore sample size needs for your particular applications.

TIME TO EVENT MODELING

Time to event modeling is used when the outcome of interest is the time until an event occurs, often referred to as the **survival time**. The focal event can be of any type but it should constitute a clearly defined qualitative change that occurs at a given point in time. Examples include retiring from a job, re-occurrence of cancer, and discharge from a hospital. The most common analytic form of time to event modeling is known as **survival analysis**. The term "survival analysis" originates from the fact that many of the early applications of the analytic tool focused on the event of mortality. Interest was in how long it takes until people in a given population dies (i.e., experiences mortality) and isolating predictors of the time until mortality. However, survival analysis has been

applied to a wide variety of events and extends beyond the concept of mortality. When the focal event is positive (e.g., getting married), the use of the term "survived" is a bit of a misnomer, namely how long people have managed to "survive" the event of getting married. However, the survival term is widely used for both negative and positive events.

The methods described in this chapter focus on the case where the event in question has not occurred for all individuals in the dataset; the study observation period concludes with some of the individuals not yet experiencing the target event, although they likely will at some point in the future. Suppose a researcher seeks to study how long women use an intrauterine device (IUD) for pregnancy protection just after the IUD was first inserted. The target event is "stopped using the IUD." The researcher might follow the women who are new-users for a two year period and notes each month over the course of the 24 months the point at which each woman stops using the IUD. Some of the women will not have stopped by the end of the observation period, so their data are said to be **right censored**; the researcher does not know when they will stop using the IUD. There are research scenarios where data are said to be **left censored**. In a longitudinal study of adolescent females, a researcher might seek to document the age at which the young women experience their first period. If a study participant has already done so at study enrollment, the exact age of onset for that person technically is unknown and it said to be left-censored. Left censoring typically is less commonly addressed in the social sciences and it is more challenging statistically to do so. My focus in this chapter is on right censoring. Henceforth, when I refer to censoring, I will always mean right censoring unless otherwise so noted.

Parenthetically, if the target event has occurred for everyone in the study by the end of the observation period, then the time-to-event analyses can use standard regression methods without adjustments for censoring. The outcome variable is the time it took for the event to occur. Another approach that avoids the use of censored-based methods is to dichotomize the outcome to 0 = the event did not occur during the study period versus 1 = the event did occur during the study period. One then applies binary regression modeling to this outcome. In this case, the outcome is crude relative to survival modeling because the results cannot be generalized outside the studied observation period and because the length of the observation period often is arbitrary. Despite this, the strategy is not unreasonable as an exploratory approach to time-to-event analytics (Allison, 2014).

Left and right censoring can occur for outcomes other than time-to-event outcomes, such as when an instrument to measure levels of lead in the blood has a minimum threshold that it is sensitive to, such as lead levels > 2 μg/dL because values less than 2 μg/dL are unknown In this case, the data produced by the instrument is left censored. Mplus can adjust for left or right censoring in any outcome but this topic is

beyond the scope of the current chapter.

Censoring is sometimes construed as a general term for someone no longer providing data in a longitudinal study. It can occur for many reasons, including traditional loss to follow-up mechanisms. Survival modeling assumes that censoring occurs due to random factors, like the study ending on a fixed date or participants being lost to follow-up because they moved away for reasons unrelated to their experiencing the target event. This assumption is called **non-informative censoring**. It is roughly analogous to the assumption of data missing at random (MAR) discussed in Chapter 25.

Survival and Hazard Functions/Curves

Two core concepts in survival analyses are those of survival functions and hazard functions. A **survival function** is defined as the probability of surviving at least to time *t* across different points in time. The **hazard function** is a rate of experiencing the event (e.g., dying) at time *t* given one has survived up to that point in time. A plot of the survival probabilities produces a **survival curve** that reflects the probability of *not* experiencing the event. It starts at 1.0 and slopes downward over time as events occur. It always ranges from 0.00 to 1.00 unless it is reported as a percent, in which case it ranges from 0 to 100. A hazard curve represents the (instantaneous) rate of experiencing the event at time *t* given one has not experienced it prior to that point. It starts at zero or some other baseline value and then can stay constant, rise or fall across time depending on risk. Survival curves provide a macro-level view of how a population holds up over time. Hazard curves focus more on the short-term dynamics of risk and reflect vulnerable time periods. They lend themselves well to modeling variables that impact the event in question.

Discrete-Time and Continuous-Time Modeling

Analytic distinctions often are made between discrete-time survival analysis and continuous-time survival analysis. **Discrete-time survival analysis** models events where time is treated as ordered categories or integers, such as days, weeks, months years or some other integer referenced time bin. **Continuous-time survival analysis** relies on continuous measures of time, thereby assuming events can occur at any infinitesimal moment. For continuous-time models, the hazard refers to an instantaneous rate, i.e., the hazard at a specific moment in time per the concept of derivatives in calculus.

When calculating hazard rates for discrete-time survival analysis, a key concept is that of a **risk set**. A risk set refers to those individuals who are at risk of experiencing the target event at each specified unit of time. Using the IUD example, suppose that 300

women are selected into the study just after they have an IUD inserted. At study outset, all 300 women are in the risk set. The target event is having the IUD removed, which is assessed every month for two years. Suppose after the first month, 10 women have had the IUD removed. The number at risk during the next assessment period is reduced to 290. At the second month, perhaps 10 additional women had the IUD removed. The risk set for the third assessment period is now reduced to 280 women. Essentially, at the end of each assessment period, the risk set is reduced by the number of people who experienced the event during the prior month. The researcher might ultimately find that by the end of the 24 months, the risk set is reduced to 100 women. In discrete-time survival models, the hazard typically is estimated by dividing the number of events during the interval by the number of individuals at risk in that interval.

There are several methods for analyzing survival curves and hazard curves. A method that evolved early in the survival analysis literature is the **Kaplan Meier method** (Kaplan & Meier, 1958). The approach is useful but it is challenging for the analysis of RETs because it does not readily incorporate continuous mediators or continuous covariates into its structure. An popular alternative is the **proportional hazards model** that can readily incorporate multiple continuous and nominal predictors. This model is my primary focus here, but I also will make use of the Kaplan-Meier method at times.

The Proportional Hazard Model for Discrete-Time Designs

The Cox proportional hazards model as applied to survival analysis uses logistic regression as its fundamental building block. Key to its application is the **proportional hazard assumption**, namely that the hazard for one predictor profile will be proportional to that for another predictor profile across time. For example, suppose the event of interest is "engaging in binge drinking" during high school and the observation period is in units of semesters. The first entry into high school is time period 0, and thereafter are semesters 1, 2, 3, 4, 5, 6, 7, and 8. Suppose the predictor variable of interest is the biological sex of students. It might be found that the odds of the event occurring for males at the end of the first semester is three times higher than the corresponding odds for females, yielding an odds ratio of 3.0. The odds themselves of getting drunk by binge drinking can then increase for both males and females across time, but the odds *ratio* at any given time point is assumed to be constant, namely 3.0. In this sense, reporting the single number, 3.0, is informative because it informs us that the odds of binge drinking for males is three times greater than that for females throughout high school. In an RET, if we study the effect of an intervention on the time to experience an event of interest, then the proportional hazard assumption presumes that the ratio of the treatment group odds to the control group odds at any given point in time is the same.

For discrete data analysis in Mplus, the data needs to be set up in a special format, albeit in wide format. Suppose there are 6 time points equally spaced by a fixed time interval. Each time point has a binary Y variable where 0 = the person has not yet experienced the target event and 1 = the person experienced the event since the last time period. Let -99 be the code for missing data. Here are score patterns for three different individuals (with labels to describe the pattern) during the six time periods:

Individual	Y1	Y2	Y3	Y4	Y5	Y6
1 (uncensored)	0	0	0	1	-99	-99
2 (lost to follow-up)	0	0	0	-99	-99	-99
3 (never experienced)	0	0	0	0	0	0

Note that once the event occurs, the remaining time periods are coded as missing (Individual 1). If someone drops out of the study and does not provide data after dropping out, the remaining time periods following dropout are coded as missing (Individual 2). Finally, there likely will be individuals who do not experience the event during the period of study and who receive zeros at each time point (Individual 3).

One can translate the Cox proportional hazard model into an SEM framework by invoking a latent variable that is often abbreviated as *f* for **frailty**. It is not a traditional latent variable but rather is analogous to the latent propensity framework for logistic regression discussed in Chapter 5. The frailty latent construct is generally thought of as a continuous variable that represents a propensity to "fail" or a propensity to experience the event we seek to prevent. The six binary variables at the different time points are typically scored 0, 1 and are often represented by the letter U, namely U1, U2, U3, U4, U5 and U6. Each of these binary variables is thought to be influenced equally by a person's frailty so the path coefficients from *f* to each U are constrained to equal one another. The treatment condition (intervention versus control, scored 1 and 0 in a dummy variable) is thought to influence *f* which ultimately translates into am impact on the various U. When we introduce mediators into the model, the idea is that the treatment condition affects *f* through its impact on each mediator which, in turn are assumed to impact *f*. Figure 16.19 presents the model for the case of two continuous mediators. Note that in this case the researchers chose not to include a direct effect of the treatment condition on *f* because they felt the intervention effect on *f* would be fully accounted for by the two mechanisms/mediators that the program sought to change. Below, I show how to test the viability of this assumption. Note also that there are not correlated disturbances between the two mediators. This implies that the primary sources of the correlation between them are (a) the treatment condition common cause and (b) the correlation

between the two baseline mediators which, in turn, are thought to influence the posttest mediators. To reduce clutter, I do not show the presumed correlational structure between the exogenous variables.

To reproduce the mathematics of the original Cox proportional hazard model, all of the loadings from *f* to U are fixed to the value of 1.0. The model is strict in form in that it assumes the impact of *f* on event occurrence is the same at all time points, a constraint that can be relaxed as I discuss later. The disturbance term for *f* is assumed to follow a logistic distribution. I describe other implied constraints below. I include two baseline covariates for the mediators. I discuss below different strategies for introducing time-invariant as well as time-varying covariates for the M→*f* links but do not do so here for the sake of pedagogy. No disturbance terms are shown for the observed U because of the inherent nature of the logistic model. See Chapters 5 and 12 for elaboration.

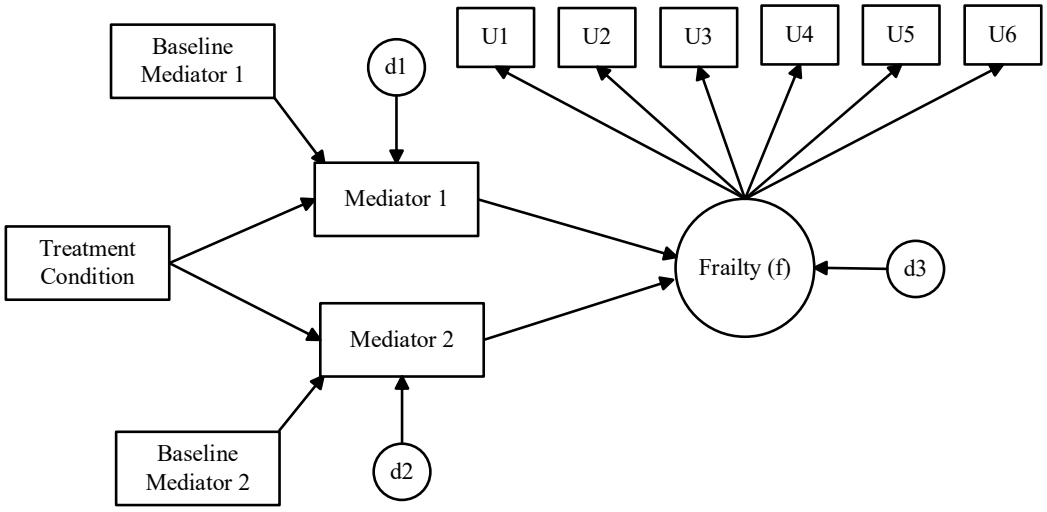


FIGURE 16.19 SEM representation of the proportional hazard discrete-time model

The example I use to illustrate the model focuses on incarcerated individuals who were about to be released from prison. The event of interest was recidivism (0 = remained out of prison, 1 = returned to prison) and the intent was to isolate hazards and "survival" times for the individuals once they were released. The study was conducted over a period of 18 months with the occurrence of recidivism recorded every 3 months, yielding a total of 6 assessments. Before release, half of the individuals participated in a program that sought to change two mediators, M1 and M2, that were thought to increase recidivism

risk. The program sought to *lower* scores on M1 and M2. The other half of the prisoners did not participate in the program and represented a treatment as usual (TAU) control group. The treatment condition was scored 0 = participated in the TAU condition, 1 = participated in the intervention condition. The mediators had a metric such that their standard deviations were close to 1.0. They were expected to independently increase fragility (*f*). The Mplus syntax appears in Table 16.14.

Table 16.14. Mplus Syntax for Proportional Hazards Model

```
1.  TITLE: Continuous time survival analysis;
2.  DATA: FILE = discreteM.txt;
3.  DEFINE:
4.    center m1b m2b (grandmean) ;
5.  VARIABLE:
6   .  NAMES ARE u1-u6 m1 m2 m1b m2b treat ;
7.    CATEGORICAL = u1-u6;
8.    MISSING ARE ALL (-9999) ;
9.    DSURVIVAL = u1-u6;
10. ANALYSIS: ESTIMATOR=MLR ;
11. MODEL:
12.   [u1$1 u2$1 u3$1 u4$1 u5$1 u6$1] (t1-t6) ;
13.   f by u1-u6@1;
14.   f on m1 m2  (p1 p2) ;
15.   f@0;
16.   m1 ON treat m1b (p4 p5) ;
17.   m2 ON treat m2b (p6 p7) ;
18.   [m1](am1) ; [m2](am2) ;
19.  MODEL INDIRECT:
20.   u1 IND treat ;
21.   u1 IND m1 ;
22.   u1 IND m2 ;
23.   MODEL CONSTRAINT:
24.   NEW (ch1-ch6 cs1-cs6 ) ;
25.   DO(#, 1, 6) ch# =  (1/(1+ exp(t#-p1*am1-p2*am2))) ;
26.   cs1 = 1-ch1 ;
27.   cs2 = cs1*(1-ch2) ;
28.   cs3 = cs2*(1-ch3) ;
29.   cs4 = cs3*(1-ch4) ;
30.   cs5 = cs4*(1-ch5) ;
31.   cs6 = cs5*(1-ch6) ;
32.   NEW (th1-th6 ts1-ts6) ;
33.   DO(#, 1, 6) th# = (1/(1+ exp(t#-p1*(am1+p4)-p2*(am2+p6)))) ;
34.   ts1 = 1-th1 ;
35.   ts2 = ts1*(1-th2) ;
36.   ts3 = ts2*(1-th3) ;
37.   ts4 = ts3*(1-th4) ;
38.   ts5 = ts4*(1-th5) ;
```

```
39.    ts6 = ts5*(1-th6) ;
40.    NEW (hdiff1-hdiff6 sdiff1-sdiff6 ) ;
41.    DO(#, 1, 6) hdiff# = th# - ch# ;
42.    DO(#, 1, 6) sdiff# = ts# - cs# ;
43.    !PLOT:
44.    !TYPE = PLOT3 ;
45. OUTPUT: SAMP STDYX CINTERVAL RESIDUAL TECH4 ;
```

Much of the syntax should be familiar. Lines 3 and 4 mean center the two continuous baseline covariates in order to make the intercepts in the equations they appear in meaningful (see Lines 16 and 17). The intercept for the equation in Line 16 will equal the control group mean for M1 and the intercept for the equation in Line 17 will equal the control group mean for M2. I assign labels to these intercepts in Line 18 (the labels are `am1` and `am2`). Line 7 declare the six U variables to be categorical. Mplus determines internally they are binary. Line 9 tells Mplus the six U variables will be part of a survival model. Line 12 estimates the thresholds for the logistic equations applied to and assigns each threshold a distinct label, namely t1 through t6. Line 13 defines the latent frailty variable and uses the `BY` command to fix the paths emanating from *f* to each U to be 1.0. Line 14 regresses *f* onto the intercept two mediators and assigns the respective path coefficients the labels `p1` and `p2`.

Line 15 fixes the residual variance of *f* to 0. This is a non-trivial decision. Doing so implies that the same value of *f* applies to all individuals (with some exceptions noted below), much like in traditional regression analyses where we assume that a regression coefficient in a linear equation applies to all individuals in the target population. An alternative formulation is to estimate the residual variance of *f*. Doing so treats *f* as a random variable that can vary from one individual to the next. I apply such a random effect model to the residual variance of *f* later to illustrate the underlying dynamic. For now, I fix it to zero because this is the standard approach of Cox discrete-time survival modeling. Some theorists argue that allowing *f* to vary across individuals is more realistic. A counterargument is that even if the values of *f* vary somewhat, perhaps they are quite close to one another such that it does not hurt to treat them as being the same for the sake of parsimony. Another reason researchers sometimes prefer a fixed variance of zero is because it is less computationally intense and more likely to converge. This is especially true when the model includes other latent variables approach and is complex in character. A third argument is that fixing the residual variance of *f* to zero often leads to greater statistical power and lower margins of error.

The statement that the fixed residual variance model assumes all individuals have the same value for *f* requires qualification. In the model in Figure 16.19, the value of *f* is by definition assumed to vary as a function of the two mediators, M1 and M2. However,

if M1 and M2 are held constant, the assumption is that there will be no further variability in *f*. The free rather than fixed residual variance model recognizes that M1 and M2 can indeed cause individual differences in frailty but it also holds that even when these variables are held constant, *f* still will show some across-individual variability due to unobserved and unmodeled sources of *f* heterogeneity. Again, the latter scenario probably is more realistic but the advantages of the fixed variance model might be compelling especially if one believes the amount of unobserved heterogeneity is small. I show later how to gain perspectives on fixing versus estimating the residual variance of *f*.

Lines 20-22 help map the mediational causal paths from the variable to the right of `IND` to the variable to the left of `IND`. I use only the binary outcome for Time 1 on the left side of `IND` because the results for the other U variables will be identical given the model setup.

Lines 23 to 42 use the `MODEL CONSTRAINT` command to calculate various supplemental statistics that assist interpretation of the model. Lines 25 to 33 calculate for the control group the hazard estimates for each time point and the survival rates. These estimates are stored in the 6 variables labeled `ch1-ch6` for the hazards and `cs1-cs6` for the survival rates. Lines 34 to 40 make comparable estimates but for the treatment group. These estimates are stored in the 6 variables labeled `th1-th6` for the hazards and `ts1-ts6` for the survival rates.

Lines 25 and 33 make use of syntax format that I have not used in this book prior to this example. It makes use of a loop shortcut to write multiple syntax statements (in this case 6 statements) using a single statement. The word `DO` tells Mplus you want to invoke the loop option. The numbers in the parentheses just after the `DO` statement include the consecutive integers Mplus will use with each pass through the loop. In this cases, there will be six passes through the loop starting with the number 1 and progressing through the number 6. The # symbol tells Mplus where to insert the loop number for the expression to the right of the parentheses. Thus, the single command will execute the equivalent of the following lines:

```
ch1 = (1/(1+ exp(t1-p1*am1-p2*am2))) ;
ch2 = (1/(1+ exp(t2-p1*am1-p2*am2))) ;
ch3 = (1/(1+ exp(t3-p1*am1-p2*am2))) ;
ch4 = (1/(1+ exp(t4-p1*am1-p2*am2))) ;
ch5 = (1/(1+ exp(t5-p1*am1-p2*am2))) ;
ch6 = (1/(1+ exp(t6-p1*am1-p2*am2))) ;
```

The expression to the right of the equal sign executes a formula for calculating a hazard estimate at each of the six time points. The predictors in this formula are all variables that have causal arrows pointing directly to the latent frailty, *f*. In this case, that

is M1 and M2. The relevant formula is from Masyn (2004). It uses the labels attached to parameters in the prior syntax. For example, `t1` is the label for the threshold value at time 1 in Line 12, `p1` is the path coefficient for M1 on Line 14, `p2` is the path coefficient for M2 on Line 14, `am1` is the covariate adjusted control group mean on M1 as reflected by the intercept on Line18 and `am2` is the covariate adjusted control group mean on M2 as reflected by the intercept on Line 18. In the expression, I calculate the hazards holding the mediators constant at their covariate adjusted mean values for the control group.

Lines 26 to 31 use the above hazards to calculate the survival estimates for each time point. Unfortunately, I am not able to apply `DO` notation for the formula given Mplus limitations so I list the six statements separately. At time 1, the survival estimate is simply one minus the hazard. At time 2, the survival estimate is one minus the hazard estimate at time 2 multiplied by the prior survivor estimate, `cs1`. At time 3, the survival estimate is one minus the hazard estimate at time 3 multiplied by the prior survivor estimate, `cs2`. At time 4, the survival estimate is one minus the hazard estimate at time 4 multiplied by the prior survivor estimate, `cs3`. And so on. The formula can be found in Newsom (2024) or any standard text on survival analysis.

In Lines 32 to 39, I repeat the process to calculate the hazard and survival estimates for the intervention group. The only twist is that the mediator means must reflect the adjusted means of the intervention group not the control group. For M1, this is accomplished by substituting `(am1+p4)` for `am1` in the expression on Line 33 and `(am2+p6)` for `am2` on Line 33. We have encountered the logic for defining group means from a dummy variable based regression in multiple prior examples throughout this book.

Lines 41 and 42 use the `DO` notation to calculate the treatment versus control differences between the hazards at each time point (Line 41) and the treatment versus control differences in the survival estimates at each time point (Line 42). I explain the remaining syntax later when reviewing the Mplus program output. [8]

I executed the syntax and first address issues of overall model fit. Indices of model fit are limited for survival modeling. The only global fit indices reported by Mplus are the model log likelihood and the classic AIC and BIC indices, neither of which are adequate standalone markers of model fit. At the local fit level, there are no modification indices nor standardized fit indices reported. Mplus reports the predicted and observed correlations between the continuous mediators M1 and M2, their baseline variables, M1B and M2B, and the treatment condition dummy variable. I can therefore examine these

[8] Mplus reports what is called the *baseline hazard* if you add the key word BASELHAZARD. The baseline hazard is the across-time hazards when the covariates predicting the latent frailty variable all equal 0. It is only meaningful if a score of 0 on the predictors is meaningful.

two submatrices to ensure there is reasonable correspondence between them. The matrices are shown in Table 16.15. There is reasonable correspondence between them.

Table 16.15. Estimated and Observed Submatrices

```
  ESTIMATED CORRELATION MATRIX

           M1             M2             M1B            M2B
           ________       ________       ________       ________

 M1           1.000
 M2           0.052          1.000
 M1B          0.496         -0.009          1.000
 M2B          0.002          0.531         -0.001          1.000
 TREAT       -0.213         -0.244          0.037         -0.010

           Correlations

           M1             M2             M1B            M2B
           ________       ________       ________       ________

 M1           1.000
 M2           0.042          1.000
 M1B          0.496         -0.052          1.000
 M2B          0.023          0.531         -0.001          1.000
 TREAT       -0.213         -0.244          0.037         -0.010
```

In the MODEL CONSTRAINT commands, I calculated the predicted hazard probabilities separately for the control and intervention conditions when the mediators equaled their respective mean values in each group (see Lines 25 and 33 in Table 16.14). There should be rough correspondence between them and the observed hazard probabilities for each group. Here is Mplus syntax for an independent analysis that uses the BASIC type of analysis to calculate the hazards in the two groups collapsing across the mediators and covariates:

```
TITLE: Hazards ;
DATA: FILE = discreteM.txt;
VARIABLE:
NAMES ARE id u1-u6 m1 m2 m1b m2b treat ;
USEVARIABLES ARE u1-u6  ;
categorical = u1-u6;
MISSING ARE ALL (-9999) ;
GROUPING IS treat (0=cntrl 1=interv);
ANALYSIS: TYPE = BASIC;
```

The proportion values reported on the output for variables U1 to U6 when the above syntax is executed are the sample observed hazards for each group. Technically, these are

conceptually different from the hazards calculated in the `MODEL CONSTRAINT` command because the latter are estimates conditional on the mediator predicted means. However, for purposes of evaluating model fit in a rough sense, the predicted and observed hazards should be reasonably close. The predicted hazards are located in the analysis that used the syntax in Table 16.14 in the section `Additional Parameters`; see `CH1` through `CH6` and `TH1` through `TH6`. in that section. Table 16.16 presents the relevant statistics. The correspondence between the predicted and observed values is reasonable.

Table 16.16. Predicted vs. Observed Hazards for the Two Conditions

	Obs Control	Pred Control	Obs Treated	Pred Treated
Hazard time 1	0.072	0.074	0.068	0.059
Hazard time 2	0.139	0.127	0.097	0.102
Hazard time 3	0.172	0.154	0.112	0.125
Hazard time 4	0.175	0.181	0.153	0.148
Hazard time 5	0.190	0.179	0.133	0.146
Hazard time 6	0.338	0.300	0.205	0.251

I can further evaluate the model by comparing it with other models (e.g., comparing the BICs for it with a random fragility model) and by generating the LISEM equivalent of modification indices. I show how to do so later.

I now turn to the three core questions for an RET, namely (1) is there a meaningful effect of the program on the outcome, (2) is there a meaningful effect of the program on the presumed mediators, and (3) is there evidence for a meaningful effect of the mediators on the outcome.

Effect of the Program on the Outcome

Treatment-outcome analyses in survival based clinical trials typically focus on hazards, survival rates, or both. I first consider hazards. There are 6 time points at which the treatment-control disparities for hazards can be evaluated. A feature of proportional discrete-time survival analysis with a latent fragility is that the log odds difference between the treatment and control groups for hazards is constrained to be equal at each of the six time points. This property allows us to capture the effects on hazards in a multivariate sense. In the Mplus syntax from Table 16.14, Line 20 was included to

provide perspectives on the treatment condition hazard effects for the binary outcome U1. Given the mathematics, the result generalizes to U2 through U6. Here is the output:

```
TOTAL, TOTAL INDIRECT, SPECIFIC INDIRECT, AND DIRECT EFFECTS

                                                    Two-Tailed
                    Estimate       S.E.  Est./S.E.    P-Value

Effects from TREAT to U1

  Total              -0.244      0.035     -6.871      0.000
```

The estimate is the log odds difference between the intervention condition minus the control condition. The difference is statistically significant (CR = 6.87, $p < 0.05$). I can express the intervention-control log odds difference using a (hazards) odds ratio by calculating the exponent of this difference. The exponent of -.244 is 0.78, indicating the odds of returning to prison are 22% lower in the intervention condition than in the control condition (i.e., 1 – 0.78 = 0.22). I can form margins of error and confidence intervals about the log odds difference by consulting the output section titled `CONFIDENCE INTERVALS OF TOTAL, TOTAL INDIRECT, SPECIFIC INDIRECT, AND DIRECT EFFECTS`: (which I edit and highlight in red the key statistics of interest):

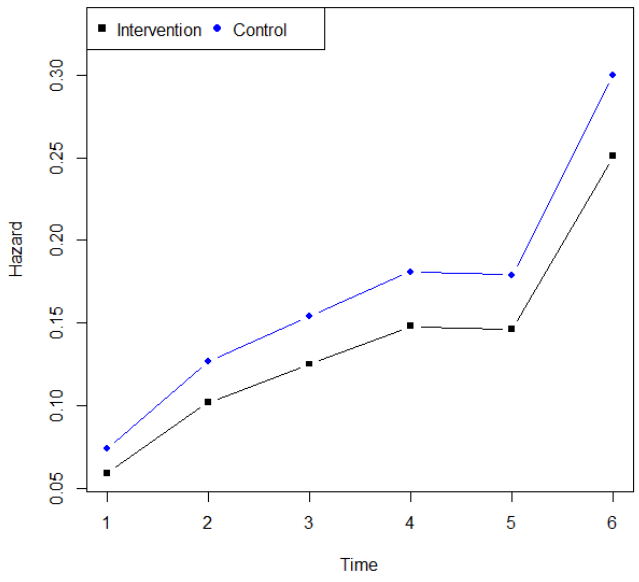
```
Effects from TREAT to U1

           Lower 2.5%    Lower 5%    Estimate    Upper 5%  Upper 2.5%
  Total       -0.313      -0.302      -0.244      -0.185      -0.174
```

The 95% confidence interval for the log odds difference of -0.244 is -.313 to -0.174. If I calculate the exponents of the lower and upper limits, I obtain the confidence interval for the hazard odds ratio of 0.78. It is 0.73 to 0.84. Note that the interval is slightly asymmetric around the 0.78 estimate. In this case, I still would be reasonably comfortable summarizing the margin of error (MOE) using the more discrepant of the two limits from the sample estimate, namely 0.78 ±0.05. Some researchers eschew reporting MOEs in such cases and are content to present the formal confidence interval.

The odds ratio approach has the nice property of capturing the group hazard disparities in a single number. However, I also like to gain a sense of the group differences in the hazards themselves not just their relative odds. I calculated the separate hazards for each time point and each group in the `MODEL CONSTRAINT` commands (Lines 25 and 33). Here are the results where the initial letter `c` represents the control condition and the letter `t` represents the treated condition; the number at the end is the time point:

	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
New/Additional Parameters				
CH1	0.074	0.009	8.655	0.000
CH2	0.127	0.011	11.107	0.000
CH3	0.154	0.013	11.532	0.000
CH4	0.181	0.015	11.727	0.000
CH5	0.179	0.017	10.563	0.000
CH6	0.300	0.022	13.634	0.000
TH1	0.059	0.007	8.309	0.000
TH2	0.102	0.010	10.453	0.000
TH3	0.125	0.011	11.033	0.000
TH4	0.148	0.013	11.045	0.000
TH5	0.146	0.014	10.139	0.000
TH6	0.251	0.020	12.767	0.000

I used the program *Temporal line plot* on the *Programs* tab of my website to plot the hazards for the two groups across time:



The hazards for the intervention group are consistently lower than for the control group as indicated by the separation between the lines. In the current context, a hazard is a conditional probability. Specifically it is the probability that the event in question (in this case, the event of returning to prison) occurs during interval *t* given that the person survived up to the start of that interval. For example, the hazard for the intervention condition at Time 4 (months 9 to 12) was 0.148. Out of those individuals who made it to the start of Time 4 without returning to prison, about 14.8 percent of them returned to

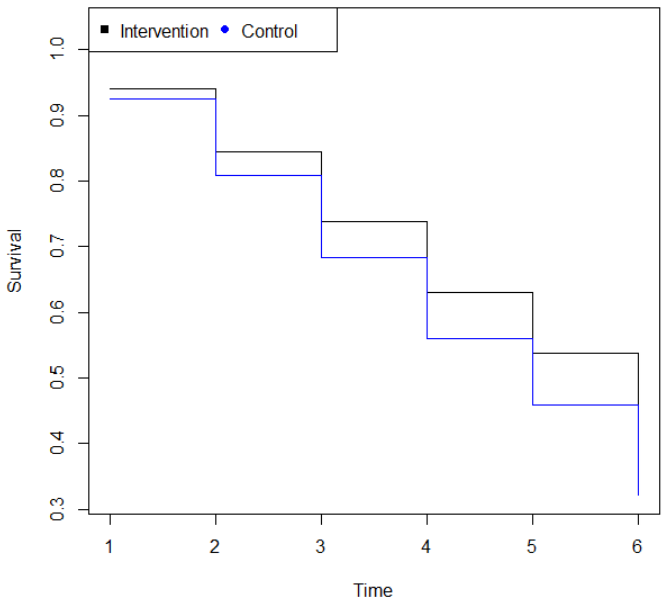
prison during the time interval spanned by Time 4. The corresponding hazard for the control condition was 0.181 or about 18.1 percent. The difference in the two hazards was 0.181-0.148 = 0.033. This is the difference in the conditional probability of the event occurring during the 9 to 12 month time period for people for whom the event had not occurred prior to the start of the interval.

Interestingly, the plot reveals that the hazard was highest in both groups at Time 6 (months 15 to 18); if individuals managed to stay out of prison until the start of the 15th to 18th month interval, the probability of them returning to prison during that interval was elevated. What is it about that period that would cause this to happen? One possibility is that many re-entry prevention programs and parole stipulations (such as intensive supervision, transitional housing, or state financial assistance) are time-limited, often lasting 3 to 12 months post-release. Perhaps these safety nets expire around Time 6 suggesting that the formerly incarcerated individuals face a sudden drop in support. This, in turn, might increase their vulnerability to relapse. Scientists typically use hazard analyses to help flag time periods during the life of the study that may be particularly problematic, such as the case for Time 6 here.

A second set of outcome indices often reported are the survival rates at each time period. I also calculated these separately for the treatment and control conditions within the `MODEL CONSTRAINT` commands. Here are the (rearranged) results where the initial letter `c` represents the control condition and the letter `t` represents the treated condition; the number at the end is the time point:

	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
New/Additional Parameters				
CS1	0.926	0.009	108.147	0.000
CS2	0.808	0.013	61.978	0.000
CS3	0.684	0.016	43.666	0.000
CS4	0.560	0.017	33.152	0.000
CS5	0.459	0.017	26.806	0.000
CS6	0.322	0.016	20.032	0.000
TS1	0.941	0.007	132.487	0.000
TS2	0.845	0.012	73.099	0.000
TS3	0.739	0.014	51.059	0.000
TS4	0.630	0.017	38.068	0.000
TS5	0.538	0.017	30.944	0.000
TS6	0.403	0.017	23.166	0.000

I used the program *Temporal line plot* on the *Programs* tab of my website to plot these survival rates for the two groups across time but now using a stair plot instead of a line plot, which is traditional for survival rates in discrete-time survival analyses:



The survival rates for the intervention group are consistently higher than for the control group as indicated by the separation between the steps at each time point. In the current context the survival rate is the probability that the event has not occurred during or before time *t*. For example at Time 4 it is an estimate of the proportion of people from the start of the study who still have not returned to prison through Month 12, the end of Time 4. Stated another way, it is the proportion of the initial group of released individuals who successfully remained in the community for the first four consecutive time intervals. A survival rate multiplied by 100 is the percentage of people who "made it" to a specific milestone (through the end of Time 4) without “failing.”

The survival rate at Time 4 was 0.560 in the control group and 0.630 in the intervention group. This means that by the 12 month mark (the end pf the period associated with Time 4), the survival percent was 63.0 – 56.0 = 7.0 percent higher in the intervention group than in the control group. Note that even though the hazard ratios charactering group disparities in hazards do not change over time, this is not the case for the survival rate differences. Survival curves typically start at study initiation at 100% in both groups but the group differences can diverge over time. A widening gap at, say, Time 4 is because the risk differences from periods 1, 2, 3, and 4 have compounded. Even if the hazard at Time 4 is identical for the groups, their survival rates can differ because the control group loses more people to “failure” in the earlier periods.

Lines 41 and 42 in the `MODEL CONSTRAINT` com60nd of Table 16.14 evaluate the hazard difference and survival differences at each time point for each of the two treatment conditions. The tests of hazard differences can yield different p values than the

odds ratio formulation because the group disparities in hazards are being characterized in different ways in the two test forms each carrying different assumptions. However, usually the p values will be in the same ball park. As an example, the hazard for the control group at Time 4 was 0.181 and for the intervention group it was 0.148. The group difference between the hazards at Time 4 (9 to 12 months) was:

	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
New/Additional Parameters				
HDIFF4	-0.033	0.005	-6.345	0.000

with the confidence intervals for the difference reported in the output section CONFIDENCE INTERVALS OF MODEL RESULTS. They are :

New/Additional Parameters	Lower 2.5%	Lower 5%	Estimate	Upper 5%	Upper 2.5%
HDIFF4	**-0.044**	-0.042	**-0.033**	-0.025	**-0.023**

Although the intervention versus control group disparities in hazards and survival rates all were statistically significant ($p < 0.05$), the question becomes whether the magnitude of the disparities are meaningful. In my experience, discussing such matters with program administrators and staff can be difficult for hazard rates because administrators and staff have trouble grasping the concept. Survival rates, on the other hand, are easier to understand. In the prison study, what is a meaningful increase in the percent of "survivors" as a function of the intervention versus control conditions? Suppose the program designers felt an increase in survivorship of at least 5% after one year (Time 4) and of at least 5% by the end of the study (Time 6) would be meaningful. All of the relevant information to evaluate the intervention relative to these meaningfulness standards are in the Additional Parameters sections linked to the MODEL CONSTRAINT commands in the CONFIDENCE INTERVALS OF MODEL RESULTS section. Here is the relevant (edited and re-ordered) information:

CONFIDENCE INTERVALS OF MODEL RESULTS

New/Additional Parameters

	Lower 2.5%	Lower 5%	Estimate	Upper 5%	Upper 2.5%
TS4	**0.597**	0.603	**0.630**	0.657	**0.662**
CS4	**0.527**	0.532	**0.560**	0.587	**0.593**
SDIFF4	**0.050**	0.053	**0.070**	0.087	**0.090**

```
    TS6                0.368       0.374       0.403       0.431       0.437
    CS6                0.290       0.295       0.322       0.348       0.353
    SDIFF6             0.058       0.062       0.081       0.100       0.104
```

At Time 4, the survival rate (translated into percents by multiplying it by 100) in the intervention condition was 63.0% ±3.3% and in the control condition it was 56.0% ±3.3%. The difference between the two survival rates was 7.0% with a 95% confidence interval of 5.0% to 9.0%. Because the lower limit of the interval is greater than or equal to the meaningfulness standard of 5%, the intervention effect is judged to be meaningful.

At Time 6, the survival rate in the intervention condition was 40.3% ±3.7% and in the control condition it was 32.2% ±3.1%. The difference between the two survival rates is 8.1% with a 95% confidence interval of 5.8% to 10.4%. Because the lower limit of the interval is greater than the meaningfulness standard of 5%, the intervention effect for Time 6 is judged to be meaningful.

In sum, the program had a meaningful effect on the survival outcome.

Effect of the Program on the Mediators

The analysis of the effects of the program on each of the continuous mediators follows the same approach from Chapter 11. Here is the relevant output:

```
MODEL RESULTS

                                                    Two-Tailed
                    Estimate       S.E.  Est./S.E.    P-Value

M1         ON
    TREAT             -0.457      0.052     -8.756      0.000
    M1B                0.504      0.026     19.107      0.000

 M2         ON
    TREAT             -0.499      0.054     -9.279      0.000
    M2B                0.542      0.027     19.881      0.000

 Intercepts
    M1                 0.465      0.037     12.613      0.000
    M2                 0.478      0.037     12.913      0.000
```

For M1, the covariate adjusted mean difference for M1 for the treatment minus control conditions was -0.46 ±0.10 (CR = 8.76, $p < 0.05$). As intended, the intervention lowered the value of M1 relative to the control group. The covariate adjusted M1 mean for the control group is the intercept for M1, which was 0.46 ±0.07 (see above). For the intervention condition, the adjusted M1 mean was -0.003 ±0.07. I calculated the latter by

re-coding the treatment dummy variable to 0 = intervention condition, 1 = control condition and then re-running the program so that the M1 intercept mow reflects the intervention group adjusted mean. From the intercept reported on the output, I could obtain its value, its estimated standard error and its margin of error.

For M2, the covariate mean difference for M1 in the treatment minus control conditions was -0.50 ±0.11 (CR = 9.28, $p < 0.05$). Also as intended, the intervention lowered the value of M2 relative to the control group. The adjusted M2 mean for the control group is the intercept for M2, 0.48 ±0.07. For the intervention condition, the adjusted M1 mean was -0.017 ±0.08.

Consultations with program staff led to setting the population meaningfulness standard to be an absolute population mean difference of 0.30 for both M1 and M2 separately. The 95% confidence interval for the M1 group difference was -0.57 to -0.35. Since the smaller of the absolute value of the two confidence limits exceeds the meaningfulness standard, the effect of the intervention on M1 is judged to be meaningful. The 95% confidence interval for the M2 group difference was -0.49 to -0.38. Since the smaller of the absolute value of the two confidence limits exceeds the meaningfulness standard, the effect of the intervention on M2 also is judged to be meaningful.

Effect of the Mediators on the Outcome

The third question focuses on the presumed effects of the mediators on the outcome. The most relevant outcomes are the binary variables (0 = did not did not "fail", 1 = "failed." At each time point. The links between M1 and M2 and the outcome thus take the form of the six logistic regressions that are the at the core of the time to event analyses. A key feature of the proportional discrete-time survival model is that the logistic coefficients for M1 and M2 are assumed to be the equal at each of the six time periods. Lines 21 and 22 in Table 16.14 isolate these coefficients for the U1 outcome and, by definition, U2 through U6. The coefficients take the form of log odds per logistic regression. Here is the (edited) output from the section labeled `TOTAL, TOTAL INDIRECT, SPECIFIC INDIRECT, AND DIRECT EFFECTS`:

```
                                                    Two-Tailed
                    Estimate       S.E.  Est./S.E.    P-Value

Effects from M1 to U1

  Total                0.239      0.046      5.206      0.000

Effects from M2 to U1

  Total                0.270      0.041      6.515      0.000
```

As expected, for every one unit that M1 increases, the log odds of "failure" increases by 0.24 ± 0.09 (CR = 5.21, $p < 0.05$). The 95% confidence interval was 0.15 to 0.33. I can calculate the exponent of this coefficient to obtain the multiplicative factor by which the odds of "failure" increases for each unit increase in M1. It equals exp(0.24) = 1.27. For every one unit that M1 increases, the odds of "failure" is predicted to increase by a multiplicative factor of 1.27 holding M2 constant. The 95% confidence interval for the multiplicative factor is exp(0.141) to exp(0.329), which equals = 1.15 to 1.39.

For every one unit that M2 increases, the log odds of "failure" is predicted to increase by 0.27 ± 0.08 (CR = 6.52, $p < 0.05$). The 95% confidence interval for the coefficient was 0.19 to 0.35. The exponent of 0.270 is 1.31. For every one unit that M2 increases, the odds of "failure" is predicted to increase by a multiplicative factor of 1.31. The 95% confidence interval for the multiplicative factor is 1.21 to 1.42.

Both M1 and M2 have standard deviations near 1.0 and thus approximate the metric of standardized scores. In criminology research, a multiplicative factor of 1.20 or greater for continuous predictors that have standardized metrics is generally considered to be meaningful. Using this standard, the effect of M2 is judged to be meaningful because the lower limit of its 95% confidence interval for the exponentiated coefficient, 1.21, is larger than the standard of 1.20. For M1, the data are more ambiguous. The lower limit of its 95% confidence interval for the exponentiated coefficient, 1.15, is less than the meaningfulness standard of 1.20. However, the upper limit, 1.39, is larger than the standard. Thus, although I can confidently conclude that the multiplicative factor is larger than 1.00, (because $p < 0.05$), I cannot confidently conclude that it is meaningful when sampling error is taken into account.

Profile Analysis for Mediator-Outcome Effects. It is possible to gain additional insights into mediator effects on survival outcomes using profile analysis. A profile is a specific multivariate pattern of scores on the predictors of a regression equation. For example, for an analysis that uses biological sex (female vs. male) and the highest grade completed (e.g., 10th grade, 12th grade) as predictors of an outcome, I might specify a profile of females who completed grade 12. Or, I might specify a profile of males who completed grade 16. Profile analyses estimate the predicted outcome for an *a priori* specified profile. This can be useful when comparing profiles against one another.

In the current example, I might want to estimate the expected survival rate for a multivariate profile of individuals who score 1 on M1 and -1 on M2, with particular interest in the survival rates for this profile at Time 4 (one year after release from prison) and at Time 6 (a year and a half after release from prison). To do so, I add the following commands to the `MODEL CONSTRAINT` section just before the `PLOT:` command in Table 16.14 (but without the line numbers):

```
1.  NEW( prof1m1 prof1m2 p1h1-p1h6 p1s1-p1s6 ) ;
2.  prof1m1=1; prof1m2=-1;  !Specify profile values
3.  !Calculate hazard at each time for profile 1
4.  DO(#, 1, 6) p1h# =  (1/ (1+ exp(t#-p1*prof1m1-p2*prof1m2) ) );
5.  !Calculate survival at each time for profile 1
6.  p1s1 = 1-p1h1 ;
7.  p1s2 = p1s1*(1-p1h2) ;
8.  p1s3 = p1s2*(1-p1h3) ;
8.  p1s4 = p1s3*(1-p1h4) ;
10. p1s5 = p1s4*(1-p1h5) ;
11. p1s6 = p1s5*(1-p1h6) ;
```

Line 2 defines the profile scores for M1 and M2. Note that if I had three mediators in the model, I would need to include values for each mediator. If I also included covariates in the model, values for them would need to be specified as well. Line 4 calculates the expected hazard for each time period using the logic of our original syntax in Table 16.14. Lines 6 through 11 translate the hazards into survival rates. Here is the output for Time 4 and Time 6.where `P1` at the beginning of the entry stands for "Profile 1":

```
                                                    Two-Tailed
                  Estimate       S.E.  Est./S.E.    P-Value

    P1S4             0.637      0.023     27.807      0.000
    P1S6             0.412      0.027     15.461      0.000
```

The estimated survival rate for individuals with the mediator profile of M1=1 and M2 = -1 at Time 4 is 0.637 or in percentage terms 63.7% ±4.6%. At Time 6 the survival rate is estimated to be 41.2% ±5.4%.

An interesting twist on profile analysis is if I define a second profile in which I hold M1 constant to the first profile value (M1 = 1) and increase M2 by one unit to M2 = 0. The calculated survival rates will then tell me the effect of increasing M2 by one unit from -1 to 0 on the survival rate while holding M1 constant at the value of 1. Here are the syntax commands I add just after the above 11 commands to calculate the profile estimates for profile 2 and then to analyze the difference between the two profiles:

```
12. NEW( prof2m1 prof2m2 p2h1-p2h6 p2s1-p2s6 ) ;
13. NEW( prof2m1 prof2m2 p2h1-p2h6 p2s1-p2s6 ) ;
14. prof2m1=1; prof2m2=0;  !Specify profile values
15. !Calculate hazard at each time for profile 2
16. DO(#, 1, 6) p2h# =  (1/ (1+ exp(t#-p1*prof2m1-p2*prof2m2) ) );
17. !Calculate survival at each time for profile 2
18. p2s1 = 1-p2h1 ;
19. p2s2 = p2s1*(1-p2h2) ;
```

```
20. p2s3 = p2s2*(1-p2h3) ;
21. p2s4 = p2s3*(1-p2h4) ;
22. p2s5 = p2s4*(1-p2h5) ;
23. p2s6 = p2s5*(1-p2h6) ;
24. !Calculate profile differences
25. NEW( p1p24 p1p26 ) ;
26. p1p24 = p1s4-p2s4 ;
27. p1p26 = p1s6-p2s6 ;
```

Here is the output for Time 4 and Time 6 for the second profile of M1=1 and M2=0:

```
                                                    Two-Tailed
                  Estimate       S.E.  Est./S.E.    P-Value

    P2S4             0.570      0.020     29.068      0.000
    P2S6             0.328      0.020     16.733      0.000
```

The estimated survival rate for individuals with the mediator profile of M1=1 and M2 = 0 at Time 4 is 0.570 or in percentage terms 57.0% ±4.0%. At Time 6, the survival rate is estimated to be 32.8% ±4.0%.

Here is the analysis of Profile 1 minus Profile 2:

```
                                                    Two-Tailed
                  Estimate       S.E.  Est./S.E.    P-Value

    P1P24            0.067      0.010      6.942      0.000
    P1P26            0.084      0.013      6.349      0.000
```

A one unit increase in M2 lowers the survival rate (expressed in percent form) by 63.7 – 57.0 = 6.7% at Time 4 and it lowers the rate by 41.2 – 32.8 = 8.4% at Time 6. By strategically constructing profiles and examining profile differences in survival rates via the `MODEL CONSTRAINT` command, one can gain interesting insights into to mediator-outcome dynamics.

Omnibus Mediation Analyses

My preference when evaluating programs is to downplay omnibus mediation effects and instead to focus on the separate links within each mediational chain of the RET model. However, Mplus reports traditional omnibus effects for the time specific binary outcomes of U1 through U6 via Lines 21 and 22 in the syntax in Table 16.14. The key results are in the output section `TOTAL, TOTAL INDIRECT, SPECIFIC INDIRECT, AND DIRECT EFFECTS.`

Here are the relevant omnibus log odd hazard differences for the treatment condition as traced through M1:

```
TOTAL, TOTAL INDIRECT, SPECIFIC INDIRECT, AND DIRECT EFFECTS

                                                    Two-Tailed
                  Estimate       S.E.  Est./S.E.    P-Value

  Specific indirect 1
    U1
    F
    M1
    TREAT             -0.109      0.025     -4.313      0.000
```

On the left, Mplus reports the mediational chain through which the treatment condition impacts U1 in the model; `TREAT` influences M1 which, in turn, influences the frailty latent variable which, in turn, influences U1. The omnibus difference in the log odds hazard through this chain is -0.109 ± 0.050 (CR = 4.31, $p < 0.05$). I can convert this value to an odds ratio by calculating the exponent of it, namely exp(-0.109) = 0.897. One minus this value is approximately 0.10, such that the hazard odds in the intervention group is about 10% lower than that for the control group.

Here are the results for M2:

```
  Specific indirect 2
    U1
    F
    M2
    TREAT             -0.135      0.026     -5.223      0.000
```

The omnibus difference in the log odds hazard through this chain is -0.135 ± 0.052 (CR = 5.22, $p < 0.05$). The corresponding odds ratio is exp(-0.135) = 0.874. One minus this value is approximately 0.13. The hazard odds in the intervention group is about 13% lower than that for the control group.

If I instead use joint significance test logic, the separate links in the mediational chains of T→M1→Y and T→M2→Y each were statistically significant, suggesting the presence of omnibus mediation for both M1 and M2.

Frailties and Unobserved Heterogeneity

The survival model for the re-entry program RET assumed that the variability of the frailty latent variable across-individuals was due to variations in the two mediators, M1 and M2. After accounting for this variability, the remaining variability in *f* was assumed to be zero, i.e., the model assumed a fixed latent frailty rather than a random one. By

contrast, a random frailty conceptualization assumes normally distributed disturbances after M1 and M2 are held constant reflecting the presence of unmeasured sources of heterogeneity, often called **unobserved heterogeneity**. Which of the two models better accounts for the data? I implement the random variance disturbance model by changing Line 15 of Table 16.14 from `f@0;` to `f*;` and then re-execute the syntax.

I can gain perspectives on the matter of relative model fit from two vantage points. First, I compare the BICs for the two models with preference given to the model with the lower BIC (see Chapter 7). For the zero disturbance variance model, the BIC on the Mplus output in the `MODEL FIT` section was 8571.42. For the random disturbance variance model it was 8571.35. The BIC difference between the models is trivial suggesting that allowing for unobserved heterogeneity in *f* has little to offer. I discuss the BIC difference approach in more detail in Chapter 7.

A second strategy sometimes used is to perform a significance test on the disturbance residual variance associated with *f* in the random disturbance variance model. Here is the relevant output from the analysis:

```
MODEL RESULTS

                                                    Two-Tailed
                  Estimate       S.E.  Est./S.E.    P-Value

 Residual Variances
    M1                0.680      0.030     22.781      0.000
    M2                0.722      0.032     22.675      0.000
    F                 0.050      0.257      0.194      0.846
```

Given a logit based framework and with factor loadings for the *f* fixed to 1.0, *f* takes on the properties of a log odds scale. The residual variance of 0.05 means that, after controlling for M1 and M2, the remaining individual-level variability in the log-odds of experiencing "failure" is 0.05, which is statistically non-significant (CR = 0.19, $p < 0.85$). The non-significant result suggests the variance of the residual frailty cannot be reliably distinguished from zero; that there is not reliable nor likely meaningful remaining frailty variance after controlling for M1 and M2.

The significance test in this case, however, comes with qualifications. The critical ratio makes the assumption that the sampling distribution of the parameter estimate is approximately normally distributed. In principle, a variance cannot be less than zero, i.e., it has a lower bound of zero. Because the null hypothesis of zero disturbance variance lies directly on the boundary of possible variance values, the underlying sampling distribution can become non-normal with inflated standard errors and low statistical

power. For this reason, most methodologists prefer the BIC difference test strategy to the significance test strategy.

In sum, the assumption of zero disturbance variance in the latent frailties does not appear to be unreasonable.

Pseudo Modification Indices

Formally, Mplus does not offer modification indices for discrete-time survival modeling to assist evaluation of model fit. As discussed in Chapter 8, there is a LISEM variant of modification indices that might be helpful. As an example for the current model, I did not include a direct effect path from the treatment condition dummy variable to the latent fragility variable. This implies that the population path coefficient is zero or functionally zero. Suppose I re-analyze the model but add the omitted path to it by changing Line 14 in Table 16.14 from

```
f on m1 m2  (p1 p2 ) ;
```

to

```
f on m1 m2 treat (p1 p2 treat) ;
```

The significance test for the path coefficient provides perspectives on whether omitting the path may have been a mistake. For the reanalysis the path coefficient was -0.112 with a critical ratio of 1.30, $p < 0.20$.

It turns out that squaring the value of the critical ratio for the coefficient will yield a value equal to or close to the modification index that occurs in traditional FISEM analyses that permit modification indices when feasible. The squared critical ratio for the above omitted path was 1.69. This index becomes my "pseudo" modification index. Note that rather than the mindless calculation of every possible modification index in a model whether they are theoretically sensical or not, the current approach allows you to focus only on localized points of stress that take into account the theoretical coherence of the omitted parameter.

A more general way of thinking about the above is that one can evaluate model fit by empirically evaluating the model's assertions of independence. In the above case, the model asserts that the association between the treatment condition and the latent *f* should be zero if I hold M1 and M2 constant. The formal way in which this independence relationship is traditionally symbolized is $f \perp\!\!\!\perp$ TREAT | M1, M2 where $\perp\!\!\!\perp$ is read as "is independent of" and | is read as "conditional on" or "holding constant."

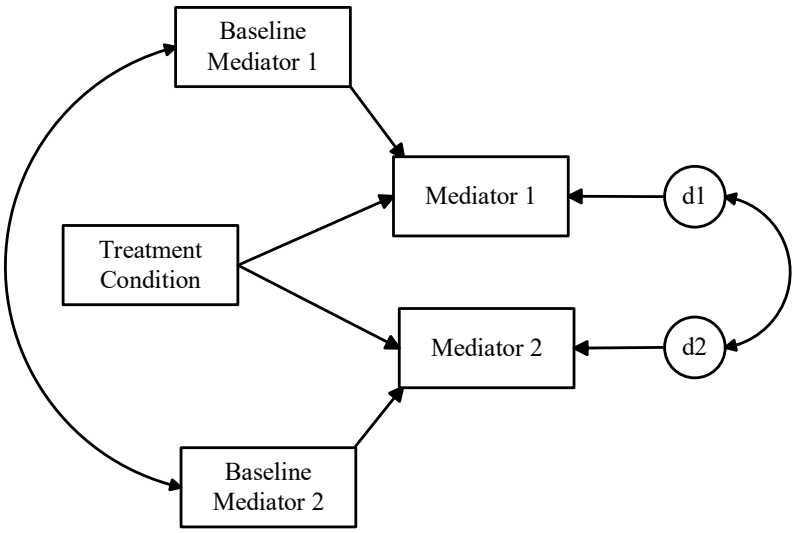
It turns out there are many independence assertions implied by prison re-entry model that can be tested using the above strategy. It sometimes is difficult for

researchers to identify the model implied independence relationships. On the *Programs* tab of my website, I provide a program called *Graph theory* that links to a website called www.dagitty.net in which you draw an influence diagram of your model and it then identifies all of the independence relations implied by the model [**program video link**]. You can then choose the key independence assertions you want to test. Corrections for multiplicity can be invoked, as desired, using the FDR method or a Holm modified Bonferroni method.

As another example of an independence assertion for the prison example is M2 ⫫ M1 | M1b, M2b, TREAT, or that M2 and M1 should be independent if I hold constant M1b, M2b, and TREAT. I evaluate this by adding a covariance between the disturbance terms for M1 and M2 via the following syntax statement

```
M1 WITH M2 ;
```

Here is the portion of the original model that shows this addition:



The correlation between d1 and d2 reflects a form of partial correlation between M1 and M2 after removing the influence of M1B, M2B and TREAT on them in accord with the specified model. The estimated correlation between disturbances for the re-analyzed data was -0.006, CR = 0.30, p < .77). The squared critical ratio was 0.09. The omission of the correlation between the disturbances seems empirically justified and is consistent with the model as originally formulated.

In sum, the formal testing of independence assertions in a model represents another form of evaluation of model-data correspondence that can be useful.

Covariates

There are multiple ways that covariates can be introduced into discrete-time survival models. As discussed earlier in this chapter, covariates can be either time-invariant or time-varying. A **time-invariant covariate** is one that does not change over time, such as a history of child trauma or past parenting practices. A **time-varying covariate** is one that potentially changes over the time periods of the survival analysis. In this book, I have reserved the term *covariate* to refer to nuisance variables that detract from making valid causal inference either in the form of confounds or inflating disturbance variance so as to reduce statistical power. However, sometimes the variables will be of substantive interest in their own right and included in the model accordingly. The current section describes ways of introducing covariates and external variables of interest into discrete-time survival modeling. To illustrate them graphically, I use a simple model with a dummy coded treatment condition T, a single mediator, M, a fragility latent variable *f*, a covariate, C, and binary outcome variables U1-U4 measured at the respective time periods of the survival analysis. I omit disturbance terms to avoid clutter.

Consider a time invariant confound, C, that I seek to statistically control when evaluating the effect of a mediator on an outcome. One way I can model the data is to add C as a predictor of each of the separate U variables, per Figure 16.20.

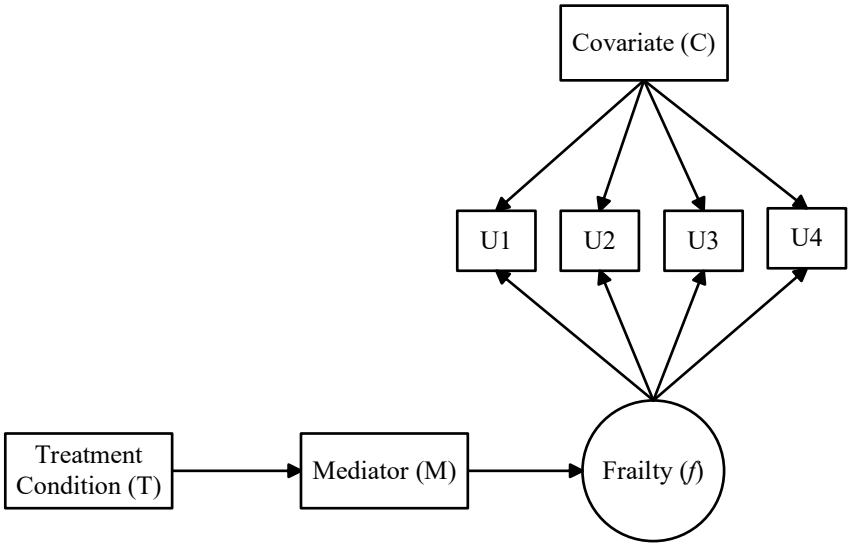


FIGURE 16.20 Time invariant covariate control via direct effects on Ul

Usually the C is mean centered to obtain estimates of the thresholds linking *f* to the different U at the average value of C.

An alternative strategy for time-invariant covariate control is to use the mean centered covariate as an additional predictor of the frailty variable per Figure 16.21.

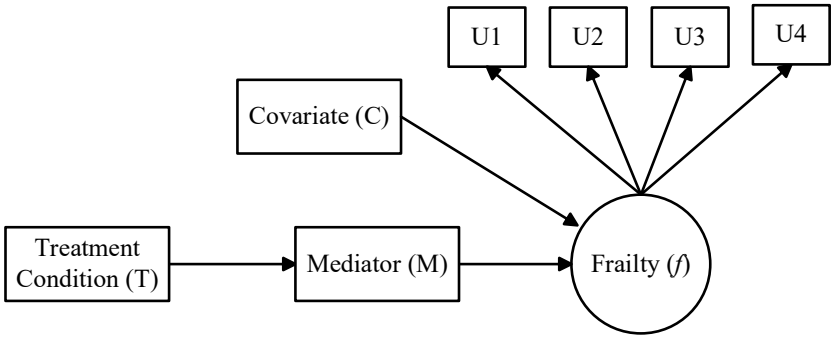


FIGURE 16.21 Covariate control via direct effect on *f*

In the latter strategy, the mathematics are such that the impact of the covariate on the various U is assumed to be equal at all four time periods. In the former strategy, the impact of the covariate on the four U is allowed to vary across the time periods. The latter strategy is more parsimonious but also can result in biased parameter estimates if the covariate impact on U changes from one period to the next.

For time-varying covariates, the traditional strategy is to include the mean-centered covariate as a synchronous predictor for each U, per Figure 16.22. The path coefficients can be constrained to be equal across time points, as one deems theoretically appropriate.

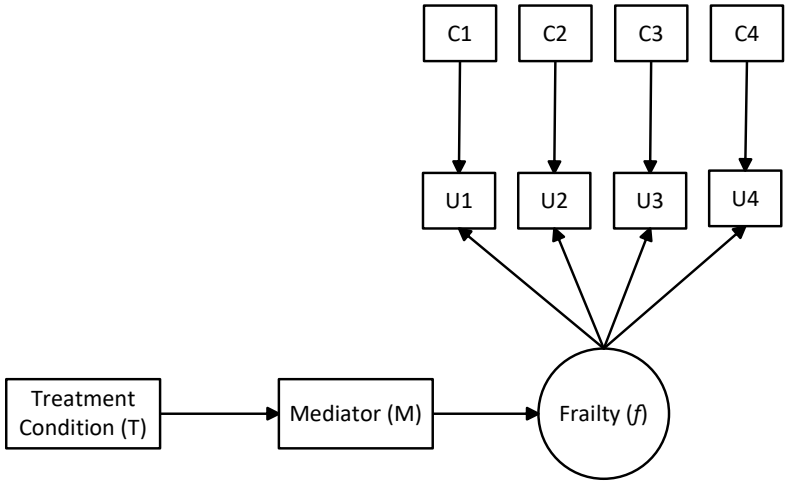


FIGURE 16.22 Time-varying covariate control

Relaxation of Constraints and Addressing Assumption Violations

The classic proportional discrete-time survival model makes multiple assumptions, the primary ones being (1) **proportionality**, which is the assumption that the effect of any predictor on the odds of experiencing the event is constant across all time periods. This assumption implies that hazard ratio between two groups is the same across time, (2) **conditional independence**, which is the assumption of independent observations after conditioning on predictors and covariates, and (3) **logit linearity,** which refers to the presumed linear relationship between the predictors and the discrete-time hazard measured on a log-odds scale.

Proportionality. The assumption of proportionality has received considerable attention and strategies for relaxing it have evolved. I illustrate key concepts using shorthand Mplus syntax focusing only on the `MODEL` , `MODEL INDIRECT` and `MODEL CONSTRAINT` commands. The initial `TITLE`, `DATA`, `DEFINE`, `VARIABLES` and `ANALYSIS` lines remain the same as does the `OUTPUT` line, unless otherwise noted. My goal is to explain the modeling approaches and the rationale behind them but not get caught up in full syntax specification and output interpretation.

One approach to nullify the proportionality constraints is to remove the latent frailty variable from the model and instead add direct paths emanating from M1 and M2 to each U. This strategy is analogous to conducting a separate logistic regression on each U separately, but doing so in a single multivariate Mplus run and with some modifications discussed later. Here is the core syntax:

```
1.  MODEL:
2.  u1 on m1 m2 (p1u1 p2u1) ;            !regress each u on m1 and m2
3.  u2 on m1 m2 (p1u2 p2u2) ;
4.  u3 on m1 m2 (p1u3 p2u3) ;
5.  u4 on m1 m2 (p1u4 p2u4) ;
6.  u5 on m1 m2 (p1u5 p2u5) ;
7.  u6 on m1 m2 (p1u6 p2u6) ;
8.  [u1$1 u2$1 u3$1 u4$1 u5$1 u6$1] (t1-t6) ; !estimate the thresholds
9.  m1 ON treat m1b (p4 p5) ;  !regress m1 on treat and its centered baseline
10. m2 ON treat m2b (p6 p7) ;  !regress m2 on treat and its centered baseline
11. [m1] (am1) ; [m2] (am2) ;      !label m1 and m2 intercepts
12. MODEL INDIRECT:
13. u1 IND treat ;               !map mediational chain from treat to each U
14. u2 IND treat ;
15. u3 IND treat ;
16. u4 IND treat ;
17. u5 IND treat ;
18. u6 IND treat ;
19. MODEL CONSTRAINT:
20. NEW (ch1-ch6 cs1-cs6 ) ;
```

```
21. !make changes in next line to accommodate proportionality removal
22. DO(#, 1, 6) ch# =  (1/  (1+ exp(t# - p1u#*am1 -  p2u#*am2 ))) ;
23. cs1 = 1 - ch1 ;
24. cs2 = cs1*(1-ch2) ;
25. cs3 = cs2*(1-ch3) ;
26. cs4 = cs3*(1-ch4) ;
27. cs5 = cs4*(1-ch5) ;
28. cs6 = cs5*(1-ch6) ;
29. NEW (th1-th6 ts1-ts6) ;
30. !make changes in next line to accommodate proportionality removal
31. DO(#, 1, 6) th# = (1/ (1+ exp(t# - p1u#*(am1+p4) - p2u#*(am2+p6)))) ;
32. ts1 = 1 - th1 ;
33. ts2 = ts1*(1-th2) ;
34. ts3 = ts2*(1-th3) ;
35. ts4 = ts3*(1-th4) ;
36. ts5 = ts4*(1-th5) ;
37. ts6 = ts5*(1-th6) ;
38. NEW (hdiff1-hdiff6 sdiff1-sdiff6 ) ;
39. DO(#, 1, 6) hdiff# = th# - ch# ;
40. DO(#, 1, 6) sdiff# = ts# - cs# ;
```

In Lines 2 to 7, I estimate the six logistic equations separately and assign unique labels to the path coefficients for M1 and M2. The first two characters in the label indicate if the path is for M1 or M2 (p1 versus p2). The second two characters of the label indicate the U variable that the path is targeting. In the `MODEL INDIRECT` command on Lines 13 to 18, I request mediational analyses from the treatment dummy variable to each of the U variables because the relevant coefficients can now differ across the U.

For Lines 22 and 31 in the `MODEL CONSTRAINT` commands, I modified the labels in the equations to accommodate the individuated labeling of the relevant path coefficients described above. Everything else remains the same.

The output reports all the estimates from the main time-discrete model but without the proportionality constraints. The effect of M1 and M2 on each of the U can differ and, given this, so can the effect of TREAT on each of the U. I discuss below issues of the multivariate structure of this analysis relative to conditional independence properties.

A variant of the above approach is to relax the proportionality assumption only partially rather than fully. To do so, I might use a piecewise approach where, for example, based on substantive theory I impose proportionality on the first three time periods and also on the second three time periods but permit the proportionality to differ between the two sets. I specify syntax that has much the same structure as above but now I impose coefficient equality constraints within each of the two piecewise sets to create partial proportionality. Here is the core syntax:

```
1.  MODEL:
2.  u1 on m1 m2 (p1u1 p2u1) ;   !regress each u on m1 and m2 but with
3.  u2 on m1 m2 (p1u1 p2u1) ;   !equality constraints via common labels
4.  u3 on m1 m2 (p1u1 p2u1) ;   !for piecewise model
5.  u4 on m1 m2 (p1u4 p2u4) ;
6.  u5 on m1 m2 (p1u4 p2u4) ;
7.  u6 on m1 m2 (p1u4 p2u4) ;
8.  [u1$1 u2$1 u3$1 u4$1 u5$1 u6$1] (t1-t6) ; !estimate the thresholds
9.  m1 ON treat m1b (p4 p5);  !regress m1 on treat and its centered baseline
10. m2 ON treat m2b (p6 p7);  !regress m2 on treat and its centered baseline
11. [m1] (am1) ; [m2] (am2) ;      !label m1 and m2 intercepts
12. MODEL CONSTRAINT:
13. NEW (ch1-ch6 cs1-cs6 ) ;
14. DO(#, 1, 3) ch# =  (1/  (1+ exp(t# - p1u1*am1 -  p2u1*am2 )) ) ;
15. DO(#, 4, 6) ch# =  (1/  (1+ exp(t# - p1u4*am1 -  p2u4*am2 )) ) ;
16. cs1 = 1 - ch1 ;
17. cs2 = cs1*(1-ch2) ;
18. cs3 = cs2*(1-ch3) ;
19. cs4 = cs3*(1-ch4) ;
20. cs5 = cs4*(1-ch5) ;
21. cs6 = cs5*(1-ch6) ;
22. NEW (th1-th6 ts1-ts6) ;
23. DO(#, 1, 3) th# = (1/ (1+ exp(t# - p1u1*(am1+p5) - p2u1*(am2+p6)) ) )  ;
24. DO(#, 4, 6) th# = (1/ (1+ exp(t# - p1u4*(am1+p5) - p2u4*(am2+p6)) ) )  ;
25. ts1 = 1 - th1 ;
26. ts2 = ts1*(1-th2) ;
27. ts3 = ts2*(1-th3) ;
28. ts4 = ts3*(1-th4) ;
29. ts5 = ts4*(1-th5) ;
30. ts6 = ts5*(1-th6) ;
31. NEW (hdiff1-hdiff6 sdiff1-sdiff6 ) ;
32. DO(#, 1, 6) hdiff# = th# - ch# ;
33. DO(#, 1, 6) sdiff# = ts# - cs# ;
34. NEW (setdiff1 setdiff2);
35. setdiff1 = p1u1 - p1u4;
36. setdiff2 = p2u1 - p2u4;
```

On Lines 2-7, note my use of common labels within each set of predictors to force proportionality on the predictors within the respective sets. In the `MODEL CONSTRAINT` command, notice I slightly altered the `DO` lines to accommodate the piecewise structure of the model. I also added Lines 33-36 to (a) test if the path coefficient for M1 in the first piecewise set is different from the path coefficient for M1 in the second piecewise set and (b) test if the path coefficient for M2 in the first piecewise set is different from the path coefficient for M2 in the second piecewise set.

I provide the Mplus outputs of both the fully-relaxed and piecewise analyses on the *Resources* tab of my webpage for Chapter 16.

In sum, researchers can fully relax the proportionality assumption or partially relax it providing a great deal of flexibility to their modeling efforts. Another strategy that I do not consider here is to apply mixture modeling to the analysis. **Mixture modeling** is a specialized statistical approach to identify unobserved (latent) subgroups within a population. Instead of assuming the data comes from a single population, mixture modeling uses a categorical latent variable to group individuals who share similar response patterns. Its application to discrete-time survival analysis is discussed in Muthén and Masy (2005), Newsom (2024), and Grimm (2026). Substantive applications of the method include Lasky et al. (2020) and Petra et al., 2011). I discuss mixture modeling in Chapter 21.

Conditional independence. The concept of local independence in discrete-time survival models refers to assumption that the probability of individuals experiencing the target event at a given time point is conditionally independent of their event status at all other time points. A practical consequence of the assumption is that the association between the various U outcomes should be zero after holding constant the model-specified determinants of them.

In the original prison re-entry example, the associations between the U across time are impacted by two continuous mediators, M1 and M2, through the frailty latent variable f. The latent frailty, in turn, equally affects the U by virtue of the fixed factor loadings of 1.0 for $f \rightarrow$U. The uniform loadings imply that the effect of the hidden vulnerability captured in f is the same throughout all six time periods. This further implies that the unconditional associations between all possible pairs of U are equal. Such dynamics can be altered in one's model by allowing, for example, unequal $f \rightarrow$U loadings but such is the structure of the model as traditionally implemented. Note that if the disturbance variance of f is fixed at zero per the prison re-entry example, then M1 and M2 are assumed to be the sole drivers of variability in f. If the disturbance variance of f is allowed to be non-zero, then f also reflects unobserved or unmeasured sources of heterogeneity in frailty. The latter model is thus broader in scope.

Tests of conditional independence for the binary U across time are not straightforward. Mplus does not report the conditionalized associations because of the mathematical complexities of doing for logistic modeling. The task is more straightforward if one uses instead probit modeling with WLSMV estimation but such a shift changes the fundamental nature of the model from one that assumes proportionality. The values of conditional associations between the U in the two model forms can be similar or they can be quite different.

Ad hoc methods for estimating the conditional association between a given pair of U have been suggested. One method involves introducing phantom latent variables (see

Muthén, 2005, 2011) but these methods are cumbersome and not feasible for models with a moderate to large number of time periods (e.g., five or greater). The approach also introduces computational complexities because they require numerical integration.

For some survival models, you can add the `TECH10` keyword on the `OUTPUT` line of Mplus and it will then provide observed and expected two-way contingency tables of event indicators for all possible pairs of time periods. In some cases, standardized residuals also are reported. However, none of these features are available for survival models with covariates, a scenario that is common for RETs.

For technical reasons, Mplus does not permit use of a `WITH` command for disturbances of the U in logistic-based survival models. As an independence diagnostic, some researchers add first-order auto-regressive paths between the U, for example U1→U2, U2→U3, U3→U4, U4→U5 and U5→U6 over and above the f→U . Statistically significant autoregressive path coefficients imply violations of conditional independence whereas non-significance is taken as evidence of assumption viability. The larger the autoregressive coefficients, the greater the violation, everything else being equal. A shortcoming of this diagnostic strategy is the frequent occurrence of model non-convergence when the autoregressive paths are introduced (which happened for the prison re-entry example). Also, the autoregressive effects only test a limited dependency structure between conditionalized U.

Yet another strategy to asses violations of conditional independence is to fit a model with the factor loadings from f to U all fixed to 1.0 and then compare the fit of the model to an alternative model that allows one or more of the factor loadings to be estimated and thus deviate from 1.0. The alternative model allows the associations between the observed U to differ across time points.[9] The models are compared using the BIC indices reported on the Mplus output. If the unconstrained loading model fits decisively better, then this suggests conditional independence in the constrained model might be violated.

In the prison re-entry model, the BIC for the fixed loading model was 8571.42. For the unconstrained loading model it was 8604.42. There was a sizeable BIC disparity between the two models but it favored the fixed loading model (because better fitting models yield smaller BICs – see Chapter 7). This result probably occurred because the BIC index rewards model parsimony. By estimating the additional factor loadings, the complexity of the unconstrained model increased non-trivially. When I simplified the unconstrained model by estimating smaller subsets of the six loadings, the BICs did not indicate significant improvement in any of them.

[9] At least one of the loadings must be fixed to 1.0 for purposes of statistical identification.

Although these results are consistent with the presence of conditional independence, there are shortcomings to the strategy that weaken such a conclusion. First, the test only provides an omnibus test of conditional independence. It does not flag where or why local independence is violated. Second, in discrete-time survival models, constraining the factor loadings to be equal forces the proportional hazard assumption onto the data. Violations of non-proportional hazards and of conditional dependence are distinct statistical issues but the above BIC difference strategy confounds the two. Third, freeing one or more of the loadings requires estimation with numerical integration which is computationally demanding and often leads to non-convergence or unstable parameter estimates. This, in turn, can render the BIC values unreliable.

In the final analysis, assessing violations of conditional independence in proportional hazards discrete time survival models that include predictors and covariates is challenging. We typically do the best we can to assess assumption viability via some combination of the above (imperfect) methods. However, since we usually cannot conclusively rule it out, the question becomes how much trouble we get into if we violate the assumption.

There have been multiple simulation studies to address the matter. Muthén and Masyn (2005) found that unmodeled local dependencies in proportional discrete-time survival models rarely create meaningful bias in the structural coefficients of the model but that they can in some cases result in downward bias of the estimated standard errors of model parameters. Oort (1998) and Garrido, Abad and Ponsoda (2016) made similar conclusions but the latter researchers emphasize that the degree of non-independence needs to be fairly strong to be consequential (e.g., $r > 0.40$). The consensus in the field seems to be that small to moderate deviations from conditional independence generally are workable but that a model's robustness depends, in part, on which parameters you focus on. As well, one must keep in mind the possible downward bias of standard errors, which can inflate Type I errors. Bottom line: Approach your results with humility.

Earlier I discussed a modeling strategy that relaxed the proportionality assumption but used a multivariate model in Mplus to simultaneously estimate model coefficients for the effects of the mediators on the binary outcomes at each time point. The matter of conditional independence among the U needs to be thought through for such applications. In survival analysis, once individuals experience the target event, they drop out of the risk set for subsequent periods. In such cases, estimating unconstrained logistic regressions for each time period using only the eligible risk set for that time period should not pose problems in terms of coefficient bias at a given time point. Technically, if the same individuals appear in the data across multiple time periods, then ignored dependencies

can intrude on standard errors if one seeks to make multivariate-like statements. Using the MLR robust estimator in Mplus helps compensate for the bias.

Linearity. As noted, the logistic-based discrete-time survival model assumes a linear relationship between continuous predictors (M1 and M2 in the prison re-entry example) and the Pon soda log odds of U at each time point. One strategy for gaining perspectives on if such linearity is viable is to plot multivariate smoothers based on a generalized additive model (GAM) discussed in Chapter 15. On the *Programs* tab of my website is a program called *Linearity diagnostics* that uses the R package *mgcv* to plot the log odds of U as a function of continuous predictors controlling for the other predictors in the target equation. For each time point in the prison re-entry example, I specified a logistic model predicting the log odds of U from M1 and M2. Figure 16.23 presents the multivariate smooths for M1 and for M2. The dashed lines are 95% confidence limits about the smooth. The plots include a "rug" at the bottom to provide a compact, one-dimensional representation of the location and distribution of data points on the X-axis. One should not expect the smooth to be perfectly linear but rather "functionally" linear. This was the case for M2 but M1 shows some deviations from linearity at the lower end of the M1 scale.

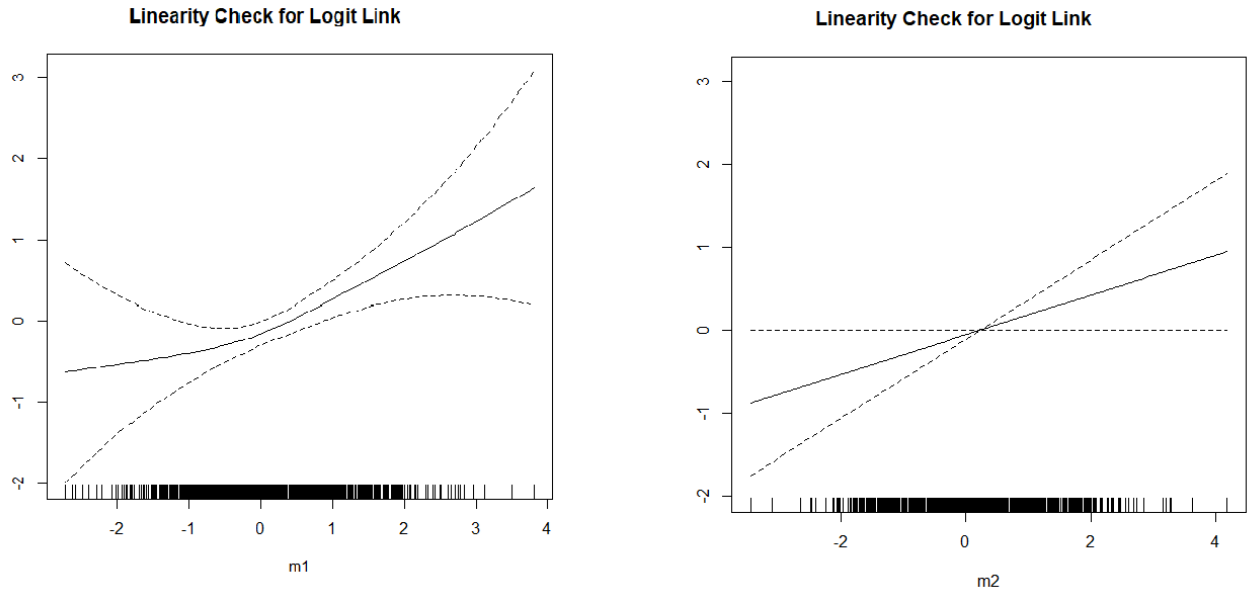


FIGURE 16.21 Linear diagnostics

In this case, I know the slight bend represents sampling error because I created simulated population data where the logit function was valid and sampled cases from it.

Of course, I would not know this in the real world as a researcher. Unless I had reasonable theory, past research, or logic that supported the bend, I would probably conclude the relationship is "linear enough." Also, when I did the same analysis at the other time points, the slight bend at the lower end of M1 did not replicate.

An alternative (or complementary) strategy is to use exploratory polynomial regression where I separately predict U at each time point from a series of successive polynomial regressions that add power terms for a target predictor to reflect a quadratic model, a cubic model, a quartic model, and a quintic model to determine if any of them suggest non-linearity of increasing complexity. The program called *Polynomial regression* on the Program tab of my web page executes the mechanics using simplified program input. Here are the results for M1 predicting U1 for logit models of increasing complexity:[10]

```
Coefficients for linear model
             Estimate Std. Error    z value Pr(>|z|)
(Intercept) -2.732787   0.142119 -19.228851 0.000000
m2           0.238733   0.118960   2.006833 0.044767
m1           0.385333   0.128155   3.006785 0.002640

Coefficients for quadratic model
             Estimate Std. Error    z value Pr(>|z|)
(Intercept) -2.831754   0.163303 -17.340524 0.000000
m2           0.242721   0.119461   2.031799 0.042174
m1           0.321319   0.126297   2.544148 0.010954
quadratic    0.103513   0.077415   1.337118 0.181184

Coefficients for cubic model
             Estimate Std. Error    z value Pr(>|z|)
(Intercept) -2.858042   0.168359 -16.975875 0.000000
m2           0.237450   0.119422   1.988326 0.046776
m1           0.407701   0.183373   2.223344 0.026193
quadratic    0.125798   0.080589   1.560982 0.118528
cubic       -0.026401   0.040361  -0.654125 0.513031

Coefficients for quartic model
             Estimate Std. Error    z value Pr(>|z|)
(Intercept) -2.871305   0.187121 -15.344669 0.000000
m2           0.237936   0.119477   1.991486 0.046428
m1           0.393841   0.201630   1.953290 0.050785
quadratic    0.150435   0.169636   0.886812 0.375180
cubic       -0.022536   0.047491  -0.474526 0.635125
quartic     -0.003724   0.022678  -0.164206 0.869569
```

[10] All predictors are mean centered to reduce possible issues with multicollinearity

```
Coefficients for quintic model
             Estimate Std. Error    z value Pr(>|z|)
(Intercept) -3.020333   0.210175 -14.370529 0.000000
m2           0.250002   0.119790   2.087008 0.036887
m1           0.753808   0.279478   2.697201 0.006993
quadratic    0.407225   0.226983   1.794078 0.072801
cubic       -0.286323   0.148922  -1.922631 0.054526
quartic     -0.053055   0.042174  -1.258002 0.208391
quintic      0.029612   0.016665   1.776877 0.075588
```

Starting with the quintic equation and working my way downward, I examine the statistical significance of the highest order term in each model to determine if it reaches statistical significance. The only equation where the highest order term attained statistical significance was the linear equation, which is consistent with the assumed linear function. This also occurred for my analyses for M2 and when I repeated this for the other U. It seems reasonable to conclude that non-linearity is a non-issue.

Continuous-Time Survival Models

In some time-to-event RETs, instead of a small to moderate number of discrete time periods, the focus is instead on many time periods to the point that time can be treated as a continuous variable in its own right. A popular method for analyzing such designs is a **proportional hazards continuous-time regression**, a semi-parametric technique for modeling the impact of multiple predictors on the time it takes for an event to occur. Technically, the method models (log) hazards that represent the instantaneous risk of an event occurring at a specific time, *t*, given that individuals have survived to that point The method's application to SEM contexts has been described by Asparouhov (2006).

Consider the following hypothetical RET which I use as my empirical example. The study focused on the time until the event of "dietary abandonment" occurred for individuals placed on a restrictive and regimented diet. Individuals were monitored daily for compliance using specialized digital monitoring methods in which participants took digital photos of their meals which were then analyzed for nutrient content with artificial intelligence tools. Participants were followed for 24 months with the expectation that most individuals would not remain on the diet by the end of the observation period. The RET randomly assigned individuals to an intervention group (score = 1) or a control group (score = 0) in the variable called T. The intervention sought to lower two continuous mediators, M1 and M2 that were thought to increase the likelihood of dietary abandonment. M1 is the perceived hassles of making changes to one's diet. M2 are the unrealistic expectations of dietary effects. The metrics of M1 and M2 ranged from -5 to +5 and each had standard deviations near 1.0. Baseline measures were obtained for the

mediators, M1b and M2b, and were used as covariates for evaluating treatment effects on them. The higher the scores on M1 and M2, the more likely it is the diet would be abandoned earlier as opposed to later.

A variable called TIME is included in the analysis to reflect the number of months that passed for each individual until the diet was deemed by a complex algorithm to have been abandoned. These scores ranged from 0 to 24 to span the two year study period but the "months until abandonment" were scored with fractions to provide more fine-grained data. If a person remained on the diet for the full two years, they were assigned a score of 24 to reflect the last known month they were event-free and still under study observation. A separate binary variable called EVENT was created to reflect if the person experienced dietary abandonment during the course of the study (0 = no, 1 = yes). Individuals with scores of 0 at the completion of the study are right censored. The data are structured in wide format as follows (shown for 5 cases):

ID	T	M1	M2	TIME	EVENT
1	1	2.41	4.12	12.4	1
2	1	1.85	3.50	24.0	0
3	0	-0.23	2.11	8.2	1
4	0	0.54	1.98	15.6	1
5	1	3.10	-.1.02	22.7	1

Figure 16.24 presents the RET structure. Note that for this study, the intervention is assumed to have a direct effect on the outcome over and above the mediators. The outcome lacks a formal error term because the model is grounded in logit modeling.

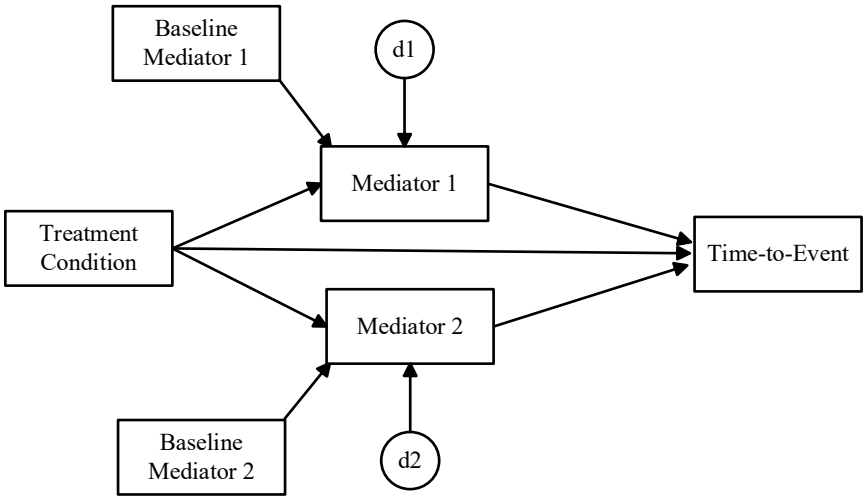


FIGURE 16.24 RET design for continuous time survival analysis

Mediational modeling for RETs with multiple mediators using continuous time survival models is not common. An alternative to the proportional hazard (PH) model considered in this chapter and that is gaining popularity is the **accelerated failure time** (AFT) model (Deboeck & Preacher, 2016; Fulcher, Tchetgen & Williams, 2017; Gelfand et al., 2016; VanderWeele, 2015). AFT models conceptualize the log of survival time as a linear function of a set of predictors. The approach requires imposing one of several parametric distributions on model disturbances with common choices being the Weibull, exponential, log-normal, log-logistic, or generalized gamma distributions. Whereas AFT strategies model the acceleration/deceleration of survival time, the PH approach models instead instantaneous risk hazards of the event occurring. AFT models focus on survival time ratios whereas PH models focus on hazard ratios. AFT models assume the distribution of survival time follows a specific parametric shape (per above), whereas the primary assumptions of PH models is that risk ratios remain constant over time. Of the two methods, the Cox-based PH model is by far the more popular one.

Some authors are skeptical of PH modeling for purposes of mediation analysis but the source of the skepticism is tied primarily to issues surrounding omnibus mediation estimation rather than the more focused link-by-link analyses I prefer (Lange & Hansen, 2011; Lapointe-Shaw et al., 2018; Martinussen et al., 2011; VanderWeele, 2011). Others raise questions about the viability of proportionality assumptions. In an article titled *The Hazards of Hazard Ratios*, Hernan (2010) notes that proportionality can be compromised by facets surrounding study design (the length of study observation horizons) and characteristics of the treatment per se (see also Shahar & Shahar, 2010). Hernan also discusses potential selection bias for hazard ratios. The bias derives from the fact that hazard ratios at a given point in time are based on individuals who have "survived" up to that point. It may be the case that time itself selects out high versus low risk individuals thereby changing the risk profile of the remaining population. As a result, the observed hazard ratio can decay over time purely because of selection processes. Others object to fact that hazard ratios, like odds ratios in logistic regression, suffer from non-collapsibility (see Chapter 12). Solutions for addressing some of these shortcomings have been proposed by Austin (2016) via the use of inverse probability of treatment weighting (IPTW). In the final analysis, if you use PH modeling, you must keep in mind the inherent limitations of the method when drawing conclusions.

There are many ways that Cox regression can be implemented in SEM contexts. I elaborate relevant issues after presenting the Mplus syntax for the numerical example, which appears in Table 16.17.

Table 16.17. Mplus Syntax for Proportional Hazards Model

```
1.  TITLE: Continuous time survival analysis;
2.  DATA: FILE = continuousM.txt;
3.  DEFINE:
4.    CENTER m1b m2b (GRANDMEAN) ;
5.  VARIABLE:
6.    NAMES = time m1 m2 T m1b m2b event;
7.    USEVARIABLES = time m1 m2 T m1b m2b event;
8.    SURVIVAL = time;   !identify the time to event variable
9.    TIMECENSORED = event (1=NOT 0=RIGHT); !identify censoring status
10. ANALYSIS: ESTIMATOR=MLR; BASEHAZARD=OFF; ! Turn off basehazard
11. MODEL:
12.   m1 ON T m1b (p1 b1);          !effect of T on mediator 1
13.   m2 ON T m2b (p2 b2);          !effect of T on mediator 2
14.   [m1] (am1) ; [m2] (am2);      !intercepts for the above
15.   time ON m1 m2 T (p3 p4 p5);   !effect of m1, m2 and T on time
16. PLOT: TYPE=PLOT3;               !generate survival plots
17. OUTPUT: Samp StdYX Cinterval Residual Tech4 ;   ! mod indices not allowed
```

Most of the syntax should be familiar. Lines 3 and 4 mean center the two nuisance covariates. Line 8 identifies the time-to-event variable and Line 9 identifies the variable indicating whether diet abandonment occurred during the study observation period (1 = so the individual is not right censored, 0 = it did not occur so there is right censoring). Line 10 turns off an analytic option in Mplus called BASEHAZARD. This option controls how the baseline hazard parameters are treated. The baseline parameters of the model describe how the risk of an event changes over time when all covariates equal zero. Activation of the BASEHAZARD option affects computational complexity as well as the estimation of standard errors. When it is turned off, the baseline hazard parameters are treated as nuisances and estimated via a method known as **profile likelihood estimation**. Mplus mathematically "profiles out" the baseline parameters by expressing them as functions of other structural parameters of the model. When turned off, the computational algorithms are much less computationally intense. The solutions tend to be more stable and convergence is more likely, especially with smaller sample size or complex models. Because they are treated as nuisance parameters, Mplus does not report them nor does it calculate their standard errors. However, point estimates of the baseline hazard parameters are calculated underneath the hood and accessible from Mplus if needed. By contrast, when BASEHAZARD is on, the baseline hazard parameters are estimated as regular parameters in the model. This allows the user to force specific parametric shapes onto the hazard function. Technical details for baseline hazard algorithms are provided in Asparouhov (2006). Asparouhov (2006) suggests that turning the BASEHAZARD option off

is typically preferred because of the statistical drawbacks of parameter bias and standard error inflation that often accompanies cases where it is turned on.

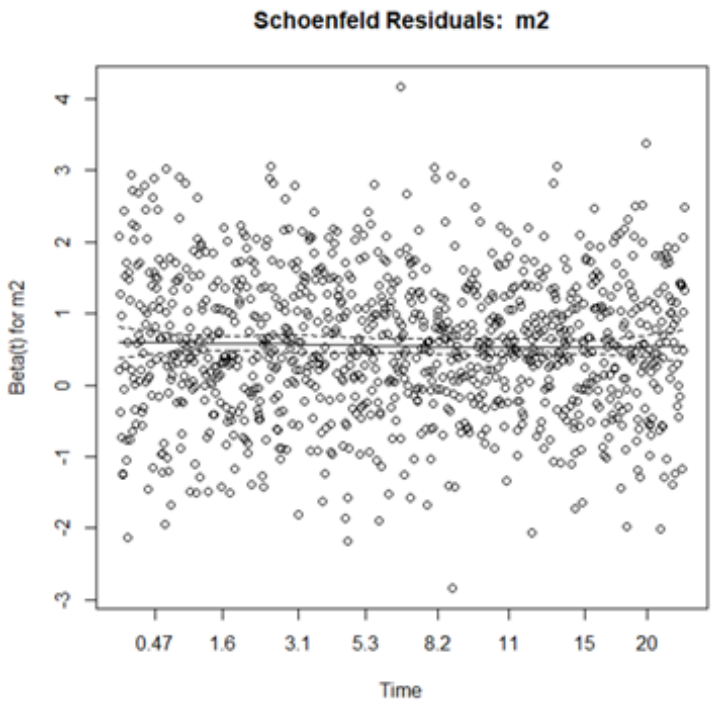
The remainder of the syntax in Table 16.17 is straightforward. I describe uses of the `PLOT` command on Line 16 below. Note that I avoided use of the `MODEL INDIRECT` command in the program because doing so results in `BASEHAZARD` being activated.

After executing the syntax, I turn to evaluating model fit. As with the discrete time survival analysis, indices of global fit are sparse. Mplus reports a log likelihood statistic and the classic information theory fit indices, the AIC and BIC. For local fit, I examine the correspondence between the estimated and observed correlation matrices for the continuous covariates and mediators. The average absolute disparity between them, which I calculated by hand, was only 0.03. The model assumes the data follow a proportional hazards pattern, which often is tested in Cox continuous-time survival regressions using **Schoenfield residuals** (Schoenfeld, 1982; Therneau & Grambsch, 2000). The test, sometimes called the **Grambsch–Therneau test**, yields a chi square statistic evaluating proportionality for each predictor as well as a global multivariate test of the predictors combined.

The test is not available in Mplus but I can gain perspectives on the matter by applying the test using continuous-time survival regression software in the R package *survival*. I did so using the program on the *Programs* tab of my website called *Survival regression* using LISEM in which I predicted the time-to-event outcome from M1, M2 and T per the causal arrows in Figure 16.24. [**program video link**]. Here are the results:

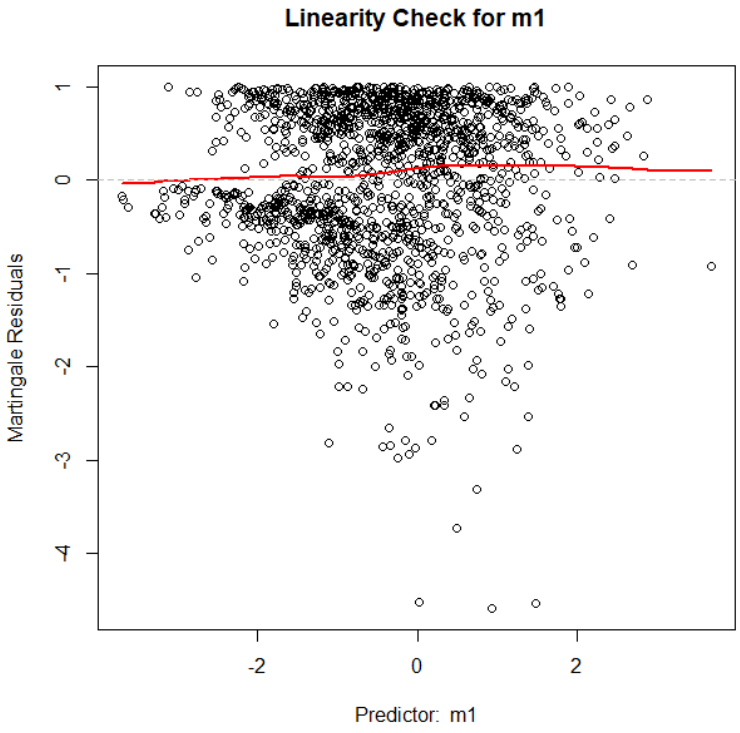
```
SCHOENFELD RESIDUALS
       chisq df    p
T      0.654  1 0.42
m1     0.035  1 0.85
m2     0.867  1 0.35
GLOBAL 1.267  3 0.74
```

None of the chi square statistics were statistically significant, which is consistent with proportionality. The program also provides Schoenfeld residual plots for each predictor in addition to a global test. It is advisable to examine these as well as the significance tests because the latter are impacted by sample size. The plots should show a flat smooth with random points about the smooth. The flat smooth indicates the coefficient for the predictor in question remained constant across time, which is a property of proportionality. See the video for the program for more information about the plots and other diagnostics [**program video link**]. As an example, here is the Schoenfeld residual plot for M2:



The dashed lines are the 95% confidence limits about the smooth. All seems to be in order. This was true of the other plots as well.

Another model diagnostic is to evaluate if the continuous predictors are indeed linear related to the log hazard as the model predicts (or presupposes). Perspectives on this matter can be addressed using **Martingale residuals**. A Martingale residual is the difference between the observed individuals' binary event status at a given time point minus the negative log of the survival probability (see Therneau & Grambsch, 2000, for details). It can range from -infinity to 1. A value near 1 means the individual "failed" very early relative to what is expected based on the model. A large negative value means they "failed" much later than the model's risk score predicted. These residuals not available in Mplus but I can again use the program called *Survival regression* on the *Programs* tab of my webpage to generate the residuals in a limited information estimation capacity. The residual plot includes a smooth that should be flat; if trend line curves non-trivially, you need to alter your modeling strategy. Here is the plot for M1:



The smooth is relative flat and consistent with the predicted linearity. This also was true of M2.

With that, I move to the three core questions for mediation analyses in RETs.

Total Effect of the Intervention on the Outcome

There are multiple strategies one can use to evaluate the overall effect of the intervention on the time-to-event outcome. As noted, I did not request for Mplus to compute the program total effect via the `MODEL INDIRECT` command because of the computational and potential statistical costs of doings so. However, I can calculate it using the `MODEL CONSTRAINT` commands by adding the following lines that make use of parameter labels just before Line 16 in Table 16.17:

```
MODEL CONSTRAINT:
  NEW(indm1 indm2 toteff);
  indm1 = p1 * p3;
  indm2 = p2 * p4;
  toteff = p5 + indm1 + indm2;
```

The line with `indm1` on the left estimates the indirect effect of the treatment condition on the outcome through M1. The line with `indm2` on the left estimates the indirect effect of the treatment condition on the outcome through M2. The line with `toteff` on the left estimates the total effect of the treatment condition on the outcome. It

sums the two indirect effects and adds to that `p5` which is the estimated effect of the treatment condition on the outcome independent of M1 and M2. Here are the results:

```
MODEL RESULTS

                                                    Two-Tailed
                  Estimate       S.E.  Est./S.E.    P-Value

New/Additional Parameters
    INDM1            -0.433      0.039    -11.056      0.000
    INDM2            -0.432      0.039    -11.150      0.000
    TOTEFF           -1.002      0.078    -12.771      0.000
```

The result for `toteff` is an intervention versus control difference on a log hazard metric. It is statistically significant and indicates the intervention condition lowered the log hazard. VanderWeele (2011) argues against exponentiating the log hazard difference for the total effect to try to make it more interpretable as a hazard ratio. His reservations are based on the fact that the two component links in the multiplication process for the indirect effects, namely T→M and M→Y, are of a different ilk metrically. Using the language of GLMs, one path is based on an identity link and the other is based on a log link. VanderWeele shows that in such cases, exponentiating the product of the coefficients can result in bias for the estimation of hazard ratios (see also Tein & MacKinnon, 2003). It is best to leave it in its log metric, which usually is not palatable to many in one's audience.

A work around is to shift to an LISEM approach via Mplus that predicts the outcome only from the treatment condition as I have done in other chapters. The syntax appears in Table 16.18.

Table 16.18. Mplus Syntax for Proportional Hazards Model

```
1.  TITLE: LISEM total effect;
2.  DATA: FILE = continuousM.txt;
3.  VARIABLE:
4.    NAMES = time m1 m2 T m1b m2b event;
5.    USEVARIABLES = time T event;
6.    SURVIVAL = time;
7.    TIMECENSORED = event (1=NOT 0=RIGHT);
8.  ANALYSIS: ESTIMATOR=MLR; BASEHAZARD=OFF;
9.  MODEL:
10.   time ON T (p);
11. MODEL CONSTRAINT:
12.   NEW(totalhr);
13.   totalhr = EXP(p);
```

```
14. PLOT: TYPE=PLOT3;       !for survival plot
15. OUTPUT: Samp StdYX Cinterval Residual Tech4 ;
```

Note that this approach does not involve the products of coefficients or summing of the direct and indirect effects. It is a more model-free approach to estimating the overall intervention effect, although it does assume the hazard ratio remains constant over time. Here are the key results:

```
MODEL RESULTS

                                                    Two-Tailed
                  Estimate       S.E.  Est./S.E.    P-Value

 TIME       ON
    T               -0.720      0.064    -11.310      0.000

New/Additional Parameters
    TOTALHR          0.487      0.031     15.708      0.000
```

The exponent of the log difference was 0.49 ±0.06, suggesting that the (instantaneous) hazard rate at any given point in time was about half the size in the intervention group than in the control group (CR on log scale = 11.31, $p < 0.05$).

The limited information estimation approach allows me to make use of resources in R that can better contextualize intervention effects. I use the program called *Survival RMST* on the *Programs* tab of my website. The program estimates intervention and control groups differences for a statistic called the **Restricted Mean Survival Time** (RMST). The RMST is the average length of time participants remain event-free up to a pre-specified time horizon, denoted as tau. In the context of the current example it is the expected number of months individuals adhere to the diet during a 24-month period, i.e., tau = 24. Mathematically, RMST is the area under the survival curve from time zero to 24. Unlike hazard ratios that express the relative risk of diet abandonment, RMST instead uses an absolute metric of time. It tells us the average number of months that individuals exhibited dietary compliance.

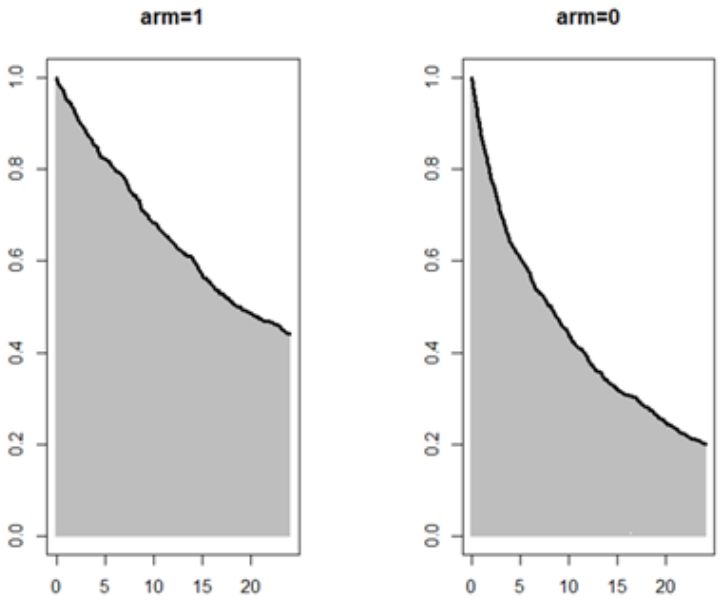
The program on my website estimates treatment and control differences in RMSTs Using the R package *survRM2*). It also reports Kaplan-Meir based survival rates across the specified time horizon using the R package *survival*.[11] Here is the output for the case where tau is 24, the entire observation period in the study (I show how to work with other values of tau shortly):

[11] The methods used in the program do not make proportionality assumption, which is a strength.

```
The truncation time: tau = 24  was specified.

Restricted Mean Survival Time (RMST) by arm
               Est.    se lower .95 upper .95
RMST (arm=1) 15.908 0.324    15.273    16.543
RMST (arm=0) 10.558 0.318     9.936    11.181

Between-group contrast
                      Est. lower .95 upper .95 p
RMST (arm=1)-(arm=0) 5.350     4.460     6.239 0

Survival Probabilities
      tau grp.n surv.prob std.err low CI high CI
arm=0  24   780    0.2000  0.0143 0.1738  0.2301
arm=1  24   720    0.4403  0.0185 0.4055  0.4781

            diff stand.err z.value p.value low CI high CI
arm1-arm0 0.2403    0.0234 10.2769       0 0.1945  0.2861

Log Rank Test (for full survival curves)
 Chi square df      p value
   131.0124  1 2.460684e-30
```

Over the 24-month study period, participants in the intervention group exhibited dietary compliance for an average of 15.91 ±0.65 months whereas those in the control exhibited such compliance or an average of 10.56 ±0.64 months. The difference is 5.35 ± 0.89 which is statistically significant ($p < 0.05$).

The program plots the survival curves for the two groups which appear as follows:



The curve drops more quickly and further in the control group compared to the intervention group. Note that the gray shaded area under the curve is larger in the intervention group than the control group. This is consistent with conceptualizing RMSTs using the concept of areas under the curve. The log rank test reported in the program is an omnibus test of the difference in the two survival curves, which was statistically significant (chi square = 131.01, df=1, $p < 0.05$).

The program also reports the survival probabilities at the end of the study observation period, namely month 24. The estimated probability of maintaining the diet until month 24 (or, stated differently, the proportion of individuals maintaining the diet until month 24) was 0.44 ±0.03 in the intervention group but only 0.20 ±0.03 in the control group. The difference between these probabilities is 0.24 ±0.05 which was statistically significant (CR = 10.28, $p< 0.05$).

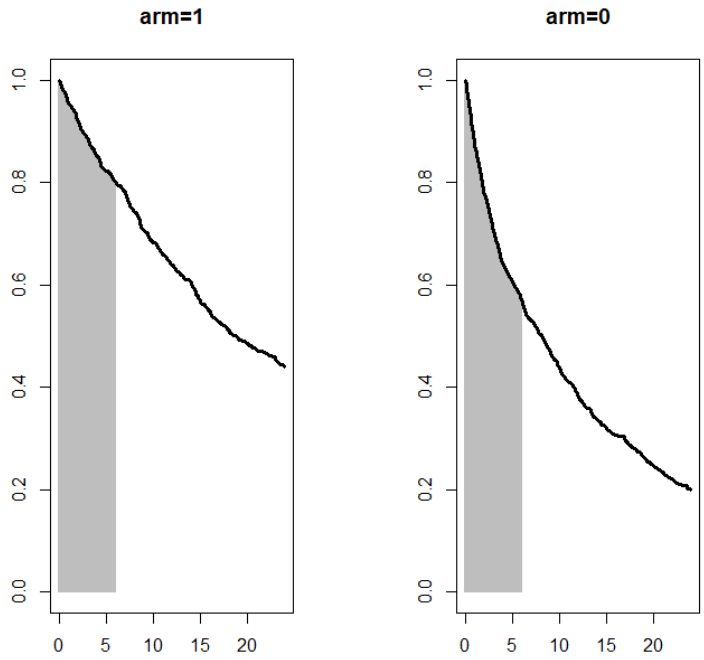
Suppose program designers and staff defined a meaningfulness standard for tau=24 to be a group difference of 2 months for RMST and a probability difference of 0.10 for survival probabilities. In both cases, the lower limits of confidence intervals for the respective statistics exceeded the meaningfulness standards so I conclude the effects of the intervention are meaningful for both of them.

In the *Survival RMST* program, it is possible to isolate the RMSTs and group survival probabilities at any point on the time dimension by changing the value of tau and re-running the program. For example, suppose I want to know the RMSTs and survival probabilities at 6 months. Here is the relevant output for a tau of 6:

```
The truncation time: tau = 6  was specified.

Restricted Mean Survival Time (RMST) by arm
              Est.    se lower .95 upper .95
RMST (arm=1) 5.326 0.058     5.213     5.439
RMST (arm=0) 4.376 0.077     4.224     4.527

Between-group contrast
                      Est. lower .95 upper .95 p
RMST (arm=1)-(arm=0) 0.950     0.762     1.139 0

Survival Probabilities (uses Kaplan-Meier)

      tau grp.n surv.prob std.err low CI high CI
arm=0   6   780    0.5692  0.0177 0.5355  0.6051
arm=1   6   720    0.8000  0.0149 0.7713  0.8298

            diff stand.err z.value p.value low CI high CI
arm1-arm0 0.2308    0.0231  9.9756       0 0.1855  0.2761
```

At the six month time point, participants in the intervention group exhibited dietary compliance for an average of 5.33 ±0.11 months whereas those in the control group exhibited compliance for an average of 4.38 ±0.15 months. The difference was 0.95 ± 0.19 months which is statistically significant ($p < 0.05$). The estimated probability of maintaining the diet until month 6 was 0.80 ±0.03 in the intervention group but only 0.57 ±0.03 in the control group. The difference between the probabilities of 0.23 ±0.05 was statistically significant (CR = 9.98, $p< 0.05$).

Here are the survival curve plots produced by the program for tau = 6.



The shaded portions of these curves are the focus of the RMST and drive home the fact that the curve to the right of the shaded portion is ignored.

Effect of the Intervention on the Mediators

For M1 and M2, higher scores promote diet abandonment. The intervention was designed to lower M1 and M2 scores, accordingly. Here is the relevant output for the T→M links that included the baseline M1b and M2b measures as covariates:

		Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
M1	ON				
T		-0.728	0.053	-13.820	0.000
M1B		0.204	0.026	7.836	0.000

```
 M2         ON
    T                 -0.775      0.054    -14.475      0.000
    M2B                0.203      0.028      7.305      0.000
```

Here are the intercepts that reflect the covariate adjusted means for the control group:

```
                                                    Two-Tailed
                   Estimate       S.E.  Est./S.E.    P-Value
Intercepts
         M1              -0.051      0.037     -1.371      0.170
         M2               0.010      0.037      0.269      0.788
```

Here are the intercepts for the reverse scored treatment dummy variable analysis, which yields the estimated covariate adjusted means for the intervention group:

```
                                                  Two-Tailed
                   Estimate       S.E.  Est./S.E.    P-Value
Intercepts
    M1               -0.779      0.038    -20.728      0.000
    M2               -0.765      0.039    -19.747      0.000
```

For M1, the adjusted posttest mean for the intervention group was -0.78 ±0.08 and for the control group it was -0.051 ±0.08. The difference between the adjusted means was -0.73 ±0.11 (CR = 13.82, $p < 0.05$). Suppose the investigators established *a priori* a meaningfulness difference standard of -0.50 or greater for M1. Both the upper and lower limits of the 95% confidence interval for the mean difference were less than this standard, suggesting the effect was meaningful.

For M2, the adjusted posttest mean for the intervention group was -0.77 ±0.08 and for the control group it was 0.01 ±0.08. The difference between the adjusted means was -0.77 ±0.11 (CR = 14.48, $p < 0.05$). Suppose the investigators established *a priori* a meaningfulness difference standard of -0.50 or greater for M2. Both the upper and lower limits of the 95% confidence interval for the mean difference were less than this standard, suggesting the effect was meaningful.

In sum, the intervention had a meaningful effect in the desired direction for both mediators.

Effect of the Mediators on the Outcome

Here is the output for estimating the presumed impact of the mediators on the outcome:

```
MODEL RESULTS

                                                    Two-Tailed
                  Estimate       S.E.  Est./S.E.    P-Value

 TIME       ON
    M1              0.595      0.032     18.374      0.000
    M2              0.557      0.032     17.306      0.000
    T              -0.137      0.071     -1.948      0.051
```

For the continuous predictors, the coefficients represent the change in the log-hazard for a one-unit increase in the target predictor holding constant the other predictors. To make the coefficients more interpretable, it is common practice to calculate the exponents of them. This isolates a multiplicative factor by which the log hazard increases (or decreases if the coefficient is negative) given a one unit increase in the predictor, holding constant the other predictors.

For M1, holding M2 and the treatment condition constant, a one-unit increase in M1 is predicted to increase the log-hazard of abandoning the diet by 0.60 ±0.06 units (CR = 18.37, $p < 0.05$). The exponent of this coefficient is 1.81. A one unit increase in M1 is associated with an 81% increase in the risk of abandoning the diet (as derived from a multiplicative factor of 1.81) at a given point in time.

For M2, holding M1 and the treatment condition constant, a one-unit increase in M2 is predicted to increase the log-hazard of abandoning the diet by 0.56 ±0.06 units (CR = 17.31, $p < 0.05$). The exponent of this coefficient 1.75. A one unit increase in M2 is associated with a 75% increase in the risk of abandoning the diet (as derived from a multiplicative factor of 1.81) at a given point in time.

For T, holding M1 and M2 constant, being in the intervention group (T=1) lowers the log-hazard of abandoning the diet by -0.14 ±0.14 (CR = 1.95, $p < 0.052$) compared to the control group (T=0). The exponent of the coefficient is 0.87 suggesting that the intervention reduces the risk of abandonment by about 13%.. Because the p value is 0.051, some researchers would declare the result as being marginally statistically significant and interpret it accordingly. Others would dismiss the effect as not having satisfied the *a priori* specified alpha level of 0.05.

A supplemental analysis that uses limited information estimation focused om RMSTs provides additional perspectives. I used the *Survival RMST* program on my website with a tau of 24 and the two mediators as mean centered covariates, Here is the output:

```
Model summary (difference of RMST)
             coef se(coef)       z     p lower .95 upper .95
intercept 12.779    0.295  43.355 0.000    12.201    13.357
arm        0.724    0.464   1.560 0.119    -0.186     1.634
m1        -3.139    0.175 -17.963 0.000    -3.481    -2.796
m2        -3.063    0.172 -17.841 0.000    -3.400    -2.727
```

For M1, the negative sign indicates that higher levels of M1 are associated with shorter average survival time for the 24 month time horizon. For every 1-unit increase in M1 on its metric, we expect a loss of 3.14 ±0.35 months of dietary compliance (CR = 17.96, $p < 0.05$) holding the other predictors in the equation constant. Suppose prior to the study, the program staff/designers felt that if a one unit decrease in M1 produced an additional month of survival time that this would signify a meaningful link between M1 and "survival." The confidence intervals suggests a meaningful link.

Higher levels of M2 also were associated with shorter average survival time for the 24 month time horizon. For every 1-unit increase in M2 on its metric, the mean survival time decreases by 3.06 ± 0.34 months (CR = 17.84, $p < 0.05$) holding the other predictors constant. Suppose prior to the study, the program staff/designers felt that a coefficient of -1.0 also would be meaningful. Both the upper and lower limit exceed this standard so the effect also is deemed meaningful.

When using RMSTs in mediation modeling, keep in mind that the path/regression coefficients are dependent on the chosen time horizon. For tau = 24, holding the treatment condition constant, a one-unit decrease in either M1 or M2 extends an individual's expected duration of dietary adherence by approximately 3.1 months over the two-year window. If I truncate the analysis to focus only on a 3-month time horizon (tau = 3), a coefficient magnitude of 3.1 would be mathematically impossible, as a one-unit change cannot shift expected adherence by more months than the window itself allows. Altering tau to 3 also changes the window of the survival experience being evaluated. While a tau of 24 captures long-term dietary patterns and late-stage survival, a tau of 3 restricts the lens entirely too early, short-term adherence. It is for this reason that methodologists emphasize the time horizon selected for RMST analysis should be driven by explicit theoretical, clinical, or behavioral milestones. When I restricted the time horizon to 3 months, the coefficient for M1 was -0.16 ±0.04 (CR = 8.46, $p < 0.05$) and for M2 it was -0.17 ±0.04 (CR = 8.82, $p < 0.05$). For every one unit that M1 or M2 decreases on their respective metrics, we expect a gain of about (0.16*30 days in a month) 5 days over the first three months (90 days) of the time horizon.

Parenthetically, as a general rule of thumb used by FDA statisticians, RMST analyses tend to be most trustworthy when the chosen tau is associated with at least 8-

12% of patients remaining at risk at that time point (Deng, 2021). Choosing a tau that is too large can result in unstable estimates.

Omnibus Mediation Analyses

There are multiple strategies one can use to evaluate the indirect effect of a mediator on the outcome. A straightforward way to test the null hypothesis of no mediation for a given mediator is to use the joint significance test as outlined in Chapter 9. In the current example, both the link T→M and the link M→Y were statistically significant ($p < 0.05$) for both mediators which leads us to reject the null hypotheses of no mediation in each case.

Earlier, I used the `MODEL CONSTRAINT` command to calculate the indirect effect of each mediator by multiplying the appropriate path coefficients in a mediational chain. I noted, however, that these estimates take the form of log hazards which are not easily . They cannot be reliably exponentiated to yield unbiased estimates of hazard ratios. Nor can their standard errors be bootstrapped in continuous-time models that depend heavily on an individual's specific "survival" time and administrative censoring point, making the original hazard structure unstable across resampled datasets (Lange, Vansteelandt & Bekaert, 2012). One strategy might be to use Monte Carlo confidence intervals as discussed in Chapter 8.

Profile Analyses

It often is helpful to conduct profile analyses to illustrate the multivariate implications of score patterns on the variables that affect the outcome. Because I used the `BASEHAZARD = OFF` framework to avoid heavy numerical integration, Mplus does not estimate an absolute baseline hazard intercept parameter. Because of this, I cannot calculate a specific hazard value for a given profile. Instead, I can use the `MODEL CONSTRAINT` command to calculate the **log-hazard relative shift** and a corresponding hazard ratio that compares two profiles in an informative way.

As an example, I might want to compare profiles of two individuals, one at low risk of diet abandonment and one at high risk of diet abandonment. I define a high risk individual as someone who did not receive the intervention (T = 0), and who has elevated M1 and M2 scores near the 80th quantiles of M1 and M2, which is M1 = 0.50 and M2 = 0.50. I define a low risk individual as someone who received the intervention (T = 1) and who had relatively low scores on M1 and M2, near their 20th quantiles, which is M1 = -1.30 and M2 = -1.30. Here are the relevant `MODEL CONSTRAINTS` commands that would be placed just before Line 16 in Table 16.17 to compare the two profiles.

```
NEW(loghi loglow contrast);
  loghi = (p3 * .5) + (p4 * .5) + (p5 * 0);
  loglow = (p3 * (-1.3)) + (p4 * (-1.3)) + (p5 * 1) ;
  contrast = loghi - loglow;
```

The statement for `loghi` calculates the log hazard for the high risk individual (ignoring the intercept) and the statement `loglow` is the log hazard for the low risk individual (ignoring the intercept). I can ignore the intercepts because in the `contrast` line where I difference the two profiles, the intercept terms cancel each other. The various p in the above syntax refer to the labels for the relevant path coefficients for M1, M2 and T. The term `EXP` calculates the exponent of the log difference to yield the hazard ratio comparison of the two profiles. Here is the output:

```
                                                    Two-Tailed
                    Estimate       S.E.  Est./S.E.    P-Value

New/Additional Parameters
    LOGHI              0.576      0.025     22.671      0.000
    LOGLOW            -1.635      0.076    -21.447      0.000
    CONTRAST           2.211      0.092     24.040      0.000
```

The difference on the log scale is statistically significantly different from 0 (CR =24.04, $p < 0.05$). The relative risk ratio between the two profiles is the exponent of 2.211 which is 9.12. This means that the hazard for the high risk individual is just over 9 times higher than the hazard for the low risk individual.

If desired, you can calculate the estimated survival probability at a given time point for a single *a priori* defined predictor profile, but doing so is best accomplished using a limited information estimation framework. See the document *Pseudo-Values in Continuous Time Survival Analysis* on the *Resources* tab of my webpage.

Concluding Comments on Continuous Time Example

To summarize the results of the analyses, the data suggest the intervention had a meaningful effect on time-to-event constructs. The instantaneous hazard rate in the intervention condition was about half the size of that of the control group. The RMST reflecting the average length of time respondents spent on the diet across the 24 month period was 15.91 ±0.65 months whereas in the control condition it was 10.56 ±0.64 months. The difference of 5.35 ±0.89 was statistically significant ($p< 0.05$) and meaningful. In addition, the survival rate at month 24 was statistically significantly ($p < 0.05$) and meaningfully different for the intervention group (0.44 ±0.03) as compared to the control group (0.20 ±0.03).

The intervention also had statistically significant and meaningful effects on the two mediators, the perceived hassles of making changes to one's diet (M1) and unrealistic expectations of dietary effects (M2). The intervention relative to the control group was estimated to decrease each mediator by about 0.75 units on their respective metrics 1.0.

Finally both mediators were statistically significantly and meaningfully related to the time-to-event outcomes. For M1, holding M2 and the treatment condition constant, a one-unit increase in M1 was associated with a 1.81 multiplicative increase in the instantaneous hazard rate; for M2, the corresponding multiplicative constant was 1.75. For the 24 month time horizon and focusing on RMSTs, for every 1-unit increase in M1 on its metric, the expected lost months of dietary compliance was about 3 months. This also was the case for M2. These represent meaningful effects.

In sum, the intervention had a meaningful overall effect on time-to-event outcomes. It also meaningfully reduced the presumed mediators it targeted and both of the mediators were meaningfully linked to the outcome. The independent effect of the intervention on the outcomes over and above the mediators was marginal at best.

Concluding Comments on RETs on Time to Event Modeling

The proportional hazard model is a popular method for analyzing continuous time-to-event data that is right censored. When its assumptions are met, it offers useful insights into such phenomena. There are a range of limited information estimation tools that can be used to supplement perspectives yielded by the SEM-based proportional hazard model. There also have been developments in survival analysis that allow these limited information tools to be used in full information estimation SEM contexts and that are worthy of your consideration. The methods rely on what are known as **pseudo-values**, a statistical approach developed by Andersen, Klein and Rosthøj (2003). The method is a data-transformation technique that converts censored time-to-event data into continuous uncensored outcomes focused on either survival probabilities or restricted mean survival times. The transformed scores can be analyzed using FISEM methods in Mplus that are much more straightforward than the proportional hazard model I described here. I provide a program called *Survival pseudo vars* on the *Programs* tab of my website that generate relevant transformed pseudo variables for both RMSTs and for survival probabilities. I discuss how to apply FISEM analyses to them using Mplus in the document titled *Pseudo-Values in Continuous Time Survival Analysis* on the *Resources* tab of my website under Chapter 16. The method is particularly useful when the assumptions of proportional hazards are untenable because pseudo-values do not make assumptions of proportionality. I urge you to look at the above document.

Traditional survival methods, like the Kaplan-Meier (KM) method and the

proportional hazard model, estimate the probability of an event assuming the event is the only possible outcome. If a participant experiences a different, competing event that makes it impossible for the primary target event to happen, traditional methods treat that person as "censored." An example of a **competing event** is a clinical trial that tracks the time until a cancer relapse occurs but where a patient dies from a heart attack before they relapse. The occurrence of the heart attack represents a competing event. Another example would be a study of time until retirement. An individual dying while in the workforce represents a competing event. When such competing events occur, it is possible that the more traditional methods overestimate survival probabilities. Statisticians have developed methods focused on **cumulative incidence functions** (CIFs) that reflect the absolute probability of a specific event occurring by a given time *t* when individuals might be exposed to competing events. The primary analytic approach to such cases is **Fine-Gray regression** (Fine & Gray, 1999). It is based on proportional hazard modeling but with mathematical adjustments. It can be applied using the R package *cmprsk*. An alternative analytic strategy for CIFs is to use the **Aalen-Johansen method** in the R package *survival*. The latter method does not make proportionality assumptions (Aalen & Johansen, 1978). A program on the *Programs* tab of my website called *Pseudo CIF vars* creates pseudo variables for CIF scenarios. If a competing event operates in your study design, the typical practice is to compare survival estimates for a traditional analysis with those in the CIF model to ensure they are comparable.

DYNAMIC STRUCTURAL EQUATION MODELING

Dynamic structural equation modeling (DSEM) is a statistical framework that integrates multilevel modeling, structural equation modeling (SEM), and time-series modeling to analyze intensive longitudinal data. I orient my discussion and illustrate the method using an RET with a binary treatment variable (scored 0 = control, 1 = intervention), two continuous mediators, M1 and M2, and a continuous outcome, Y. Suppose Y is a measure of sleep quality on a given night that is obtained on each day for 28 consecutive days, post-intervention. M1 is a measure of psychological detachment from work (PDW) on a given day, reflecting the extent to which the person was able to leave the stresses and demands of work at the office. M2 is a measure of the extent to which the individual engages in appropriate wind-down behaviors relative to sleep on a given day. These include such sleep hygienic behaviors (SHBs) as turning off all electronic devices (phones, laptops, TVs) at least 30 to 60 minutes before going to bed, going to bed about the same time each night, engaging in short low-stimulation activities

before going to bed, such as reading a book or stretching, creating a dark, quiet, and cool environment for sleeping, and avoiding late-day stimulants, such as coffee.

The measures of M1, M2 and Y were obtained on each of the 28 days. More stable multi-item baseline measures of M1 and M2 were obtained to use as time invariant covariates for the respective M1, M2 as well as a key time invariant covariate for all three variables that I call COV1. The metrics of M1, M2 and Y all were such that their approximate standard deviations were 1.5. The intervention sought to teach participants techniques for separating work life from home life (thereby addressing M1) and helping them understand and see the importance of engaging in sleep hygienic behaviors (thereby addressing M2). Control individuals received materials on an unrelated topic.

Figure 16.25 presents the two model forms that were tested via the DSEM framework, (1) a between-person model focused on person-level mean scores calculated across the 28 days, and (2) a within-person model focused on the causal impact of M1 and M2 on Y within individuals but across time. I discussed these two model forms earlier in this chapter in the section *Between-Person and Within-Person Analyses*, which you may want to review before proceeding. I omit covariates and disturbance terms from the Figure to avoid clutter but are included in the Mplus syntax.

Consider first the within-person model. This model evaluates the estimated effect of the two mediators on the outcome on a within-individual basis. A positive path coefficient *p* for M1 means on days when a person's M1 is higher than their usual average across days (holding M2 constant), their Y also is higher. Stated somewhat differently, a one unit increase in M1 from one day to the next is associated with a *p* unit increase in Y, holding M2 constant. Note that because these coefficients are estimated within-individuals, all time invariant confounds, measured or unmeasured, are controlled for. This includes variables like biological sex, gender, past experiences during childhood, genetic make-up, and so on. The coefficients are assumed to be functionally the same for all individuals, but this assumption can be loosened as discussed later. This portion of the model provides key insights into the M→Y links.

Note that the model includes a first order autoregressive effect with lag 1 for each of the variables. Doing so controls for lingering or carry over effects of a variable from the prior day to the current day. For example, for M2, it was felt that performing sleep hygienic behaviors on day *t*-1 increases the likelihood of performing those behaviors on day *t* via habit momentum processes. This effect is thought to be transitory and can decay if the routine is broken. Similarly, the effect of failing to psychologically detach from work on a given day can have carry-over or lingering effects on doing so the next day. The inclusion of these lagged effects controls for score dependencies across time in

accord with a simplex model. I discuss below how to control for such dependencies using strategies other than autoregressive effects on the target variables per se.

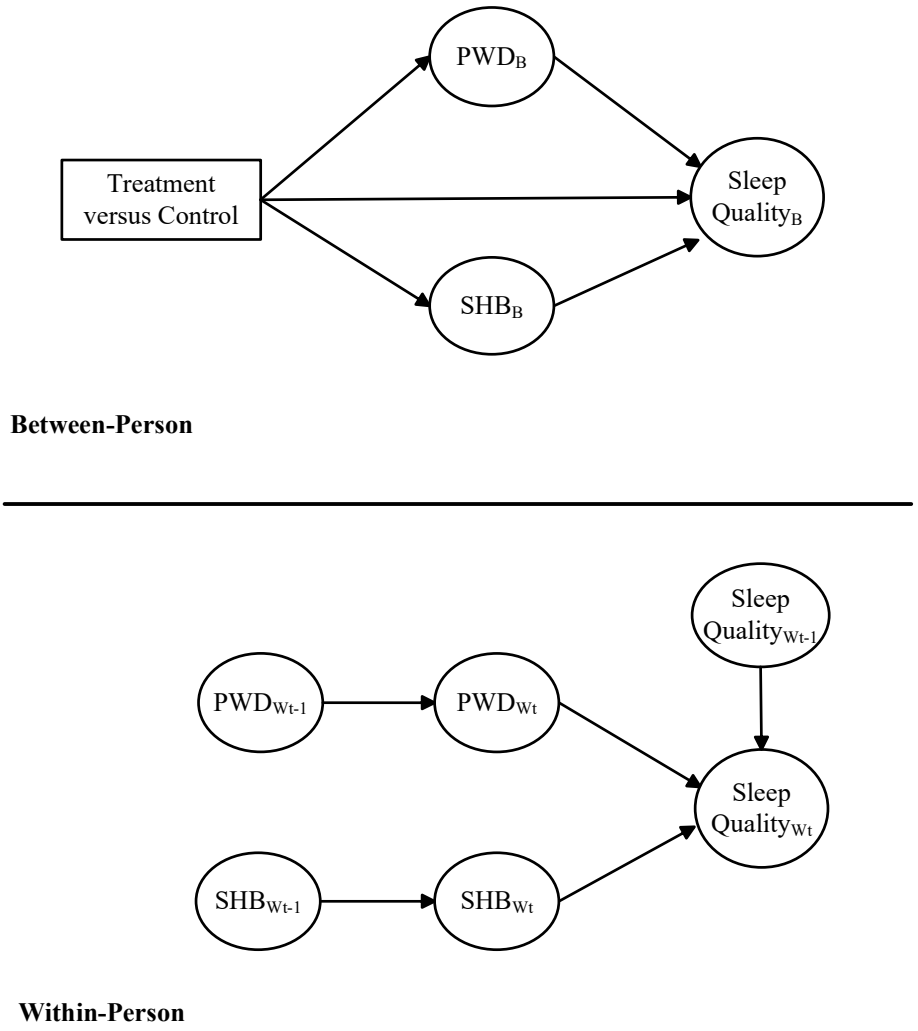


FIGURE 16.25 RET design for continuous time survival analysis

The between-person model works with the average values of M1, M2 and Y for each person across the 28 days. For M1, it posits that individuals who have higher average M1 values across days will have higher average Y values than individuals with lower M1 scores, holding M2 (and the other covariates constant). This model includes the measured time invariant variables, most notably the treatment condition, because they are, by definition, between-person in character; they do not vary across time. The path

coefficient for M1 regressed onto the binary treatment condition estimates the effect of treatment on the person's average level of M1 aggregated across the 28 days. A positive coefficient means that, on average, individuals in the intervention group had a M1 mean higher than those in the control group holding the covariates in the equation constant. This interpretation also applies to M2. A positive coefficient for when Y is regressed onto M1 indicates that individuals who have a higher average level of M1 across time also tend to have a higher average level of Y compared to individuals with lower average levels of M1 across time, holding M2 and the treatment condition constant. Similar interpretations apply to M2 and to the direct effect of the treatment condition on Y independent of M1 and M2.

Figure 16.25 uses a mix of rectangles and circles to represent variables. Influence diagrams typically use rectangles for observed variables and circles for latent variables. I retain this tradition because of a special feature of Bayesian-based DSEM. As discussed earlier in this chapter, the total variability of any variable in a longitudinal multilevel model is an additive function of its within-person variability and its between-person variability. In the current study for the outcome Y, for example,

$$var_{TOTAL}(Y) = var_{BETWEN}(Y) + var_{WITHIN}(Y)$$

where $var_{BETWEEN}$ is an index of the between-person variance of a variable and var_{WITHIN} is an index of the within-person variance of that variable, expressed using sample notation. Stated another way, the total variability of Y across all data points is a function of how the average Y for an individual varies across individuals and also how much Y varies within each person across time. The same is true for M1 and M2.

Note that for a time invariant variable like the treatment condition, there is no within-cluster variability; its variability is completely determined by $var_{BETWEEN}$. But for variables like Y, M1 and M2, both $var_{BETWEEN}$ and var_{WITHIN} contribute to their overall variability. Consider the variable Y, which has both between-person and within-person variability. In traditional multilevel modeling, the mean of Y is calculated across the 28 days to represent the between-person Y mean for a given individual. A property of this practice is that average Y across days is an *estimate* of the true between-person mean for the individual. This estimate likely is subject to sampling error which introduces a level of "noise" or uncertainty into the mean value for the person. An elegant feature of Bayesian-based DSEM as implemented by Mplus is that it explicitly takes such sampling error into account when estimating between-person and within-person variability and when calculating the Bayesian equivalent of standard errors for purposes of interval estimation and significance tests. The strategy Mplus uses is based on a complex decomposition procedure that defines latent variables for each between-within variable in

the model that are adjusted for the sampling error (see Asparouhov & Muthén, 2019). This is why I represent these variables with circles rather than rectangles. Technically, they are latent variables that have been adjusted for sampling error.

Another notable feature of Figure 16.25 is that I only include autoregressive effects at the within-person level not the between-person level. This is because autoregressive effects are inherently within-person focused: Does felt work stress from yesterday impact felt work stress for today? By contrast, between-person constructs reflect differences between people averaged over time. If person A has a higher overall 28-day average of detachment from work than person B, it essentially represents a time invariant construct collapsed over days. A person's 28-day mean cannot be "lagged" by one day.

Parenthetically, for DSEM modeling, the data needs to be structured in long rather than wide format (see the section in this chapter titled *The Basics* in the section *SEM-based Fixed Effects Panel Models*). With that, let's turn to the Mplus syntax which appears in Table 16.19.

Table 16.19. Mplus Syntax for DSEM Example

```
1.  DSEM Fixed-Coefficient Analysis;
2.  DATA:  FILE = fixedm.dat;
3.  DEFINE:
4.    CENTER cov1 m1b m2b (GRANDMEAN) ;
5.  VARIABLE:
6.    NAMES = y m1 m2 m1b m2b civ1 treat id time lm1 lm2 ly ;
7.    USEVARIABLES = y m1 m2 m1b m2b cov1 treat ;
8.    CLUSTER = id ;
9.    TINTERVAL = TIME(1);
10.   LAGGED = y(1) m1(1) m2(1);
11.   BETWEEN = treat m1b m2b cov1 ;
12.   MISSING ALL (-9999) ;
13. ANALYSIS:
14.   TYPE = TWOLEVEL ;
15.   ESTIMATOR = Bayes;
16.   BITERATIONS = 10000 (100000); BCONVERGENCE =.05 ; CHAINS = 2 ;
17. MODEL:
18.   %WITHIN%
19.    y ON m1 m2;
20.    m1 ON m1&1;
21.    m2 ON m2&1;
22.    y ON y&1;
23.    y; m1; m2;
24.    m1 WITH m2;
25.  %BETWEEN%
26.    m1 ON treat m1b cov1 (p1 b1 b2) ;
27.    m2 ON treat m2b cov1 (p2 b3 b4) ;
```

```
28.     y ON m1 m2 treat cov1 (p3 p4 p5 b5);
29.     m1 WITH m2 ;
30.     y; m1; m2;
31.    [y](ay); [m1](am1); [m2](am2);
32. MODEL CONSTRAINT:
33.    NEW (indm1 indm2 toteff) ;
34.    indm1 = p1*p3 ;
35.    indm2 = p2*p4 ;
36.    toteff = indm1 + indm2 + p5 ;
37.    NEW (m1meanc m1meant m2meanc m2meant ymeanc ymeant) ;
38.    m1meanc = am1 + p1*0 ;
39.    m1meant = am1 + p1*1 ;
40.    m2meanc = am2 + p2*0 ;
41.    m2meant = am2 + p2*1 ;
42.    ymeanc = ay + p3*m1meanc + p4*m2meanc  + p5*0 ;
43.    ymeant = ay + p3*m1meant + p4*m2meant + p5*1 ;
44. PLOT: TYPE = PLOT3; FACTORS = ALL (500);
45. SAVEDATA:  FILE = datawithlags.dat;
46. OUTPUT: STDYX RESIDUAL CINTERVAL(HPD) TECH4 TECH8;
```

Most of the syntax should be familiar. The analysis uses Mplus' multilevel framework that requires the specification of a cluster variable (see Chapter 25). In the current case, time is nested within individuals, so individuals are treated as "clusters." In Line 8. I use the unique ID identifier for each individual as the cluster identification variable, Lines 3 ad 4 mean center the three time-invariant covariates.

Line 9 identifies the variable that contains the different time periods in the analysis, which in this case is `TIME`. In my raw dataset, the variable `TIME` consists of integers for each person ranging from 1 to 28. Internally, Mplus handles the time variable by creating a time grid that orders the raw values from lowest to highest and sequentially numbers them into equal "bins" based on the unit value specified in the parentheses. For instance, if I specify `TINTERVAL = TIME(1)`, Mplus creates 28 distinct time points or bins with an interval width of 1 unit between each one. If I instead specify `TINTERVAL = TIME(.5)`, Mplus divides the timeline into narrower segments, creating 56 distinct time points with intervals of 0.5 units. Conversely, specifying `TINTERVAL = TIME(2)` collapses the grid into 14 distinct time points, with each bin spanning an interval of 2 units.

Line 10 identifies which variables Mplus should use to create internal lagged predictors for modeling purposes, meaning you do not have to manually derive these lag variables in your dataset beforehand. Specifically, Line 10 instructs Mplus to create a lag-1 version for the variables *y*, *m1*, and *m2*, where the number in the parentheses dictates the number of grid steps—expressed strictly as integers—that the software should look backward. A value of 1 signifies a first-order lag, a value of 2 represents a second-order lag, and so on. It is crucial to note that the `LAGGED` command always counts in internal

grid steps rather than raw time values. Therefore, even if I had specified `TINTERVAL = TIME(.5)` to generate 56 time points, I would still write y(1) on the `LAGGED` command to obtain a first-order lag, as writing y(.5) would result in an syntax error.

Line 11 identifies variables that are only to appear in the between-person portion of the model. They typically are time-invariant variables. If I want to restrict variables to only appear in the within-person portion of the model, I would have a corresponding `WITHIN` line with the names of those variables after the keyword.[12] If a variable is not listed on either of these lines, it is assumed by Mplus to occur in both the between-person and within-person models in one form or another. Mplus applies the aforementioned decomposition latent variable method to each of the unnamed variables.

Lines 13 to 15 tell Mplus to use a two level, multilevel model with Bayes estimation. By default, Mplus uses non-informative priors that are described toward the end of the output. Line 16 overrides the Mplus defaults for the iteration process. These values are conservative and dramatically slow processing time. I usually comment out this line and use Mplus defaults until I am sure my syntax is correct, after which I insert Line 16. For `BITERATIONS`, the first number forces Mplus to run *at least* 10,000 iterations of which the first half are discarded as burn-in. The second number allows up to 100,000 iterations if the model struggles to reach stability. The `CHAINS` option requests at least two chains. Mplus compares variation between the chains to determine if they have found the same solution. `BCONVERGENCE` sets the Potential Scale Reduction (PSR) criterion. A value of 0.05 means Mplus will only stop iterating when the PSR drops below 1.05.

Lines 17 to 23 specify the within-person model and Lines 25 to 30 specify the between-person model. The former is led by a `%WITHIN%` statement that tells Mplus the within-person model is forthcoming and the latter is led by a `%BETWEN%` statement that tells Mplus the between-person model is forthcoming. The models are specified using standard Mplus conventions. Note my use of the lagged variables in the `WITHIN` model.

Line 31 requests intercepts for the equations and labels them. Lines 32 to 36 calculate omnibus mediation effects and the total effect of the treatment on the outcome. The `MODEL INDIRECT` command is typically used to obtain these estimates but it is not allowed with Bayes. I instead calculate them using the `MODEL CONSTRAINT` command. Lines 37 to 43 calculate the model implied posttest means for the intervention and control groups for the outcome and mediators. Line 44 generates useful plots to evaluate the quality of the Bayes estimation process. Line 45 saves data from the analysis to the named file. I explain the purpose of doing so later. The `OUTPUT` line disallows modification indices and descriptive statistics; keywords for them are omitted. The HPD

[12] Mplus by default places lagged variables in the within-person model. You do not list them on the WITHIN line.

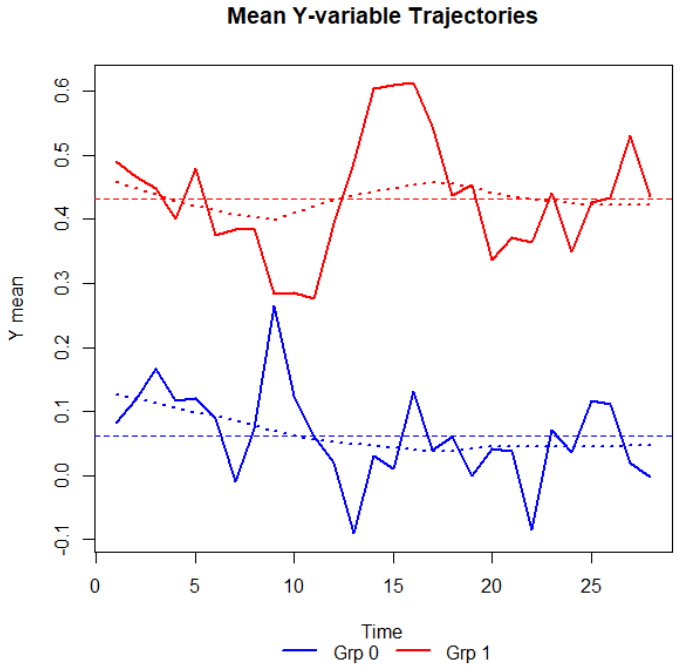
variant of credible intervals are requested and TECH8 is added to evaluate convergence.

Before discussing results, there is a technicality I should note. In what follows, I refer to the parameters that track point estimates of group central tendencies as model-implied means, sometimes more simply just as means. Technically, the values are the median of the Bayes estimated posterior distributions for means estimated under the constraints imposed by the posited SEM model and Bayesian algorithms. Such also is the case for the path coefficients. See Chapter 8 for details. If I work with traditional model-free means calculated directly on the data, I refer to them as raw means.

Model Fit

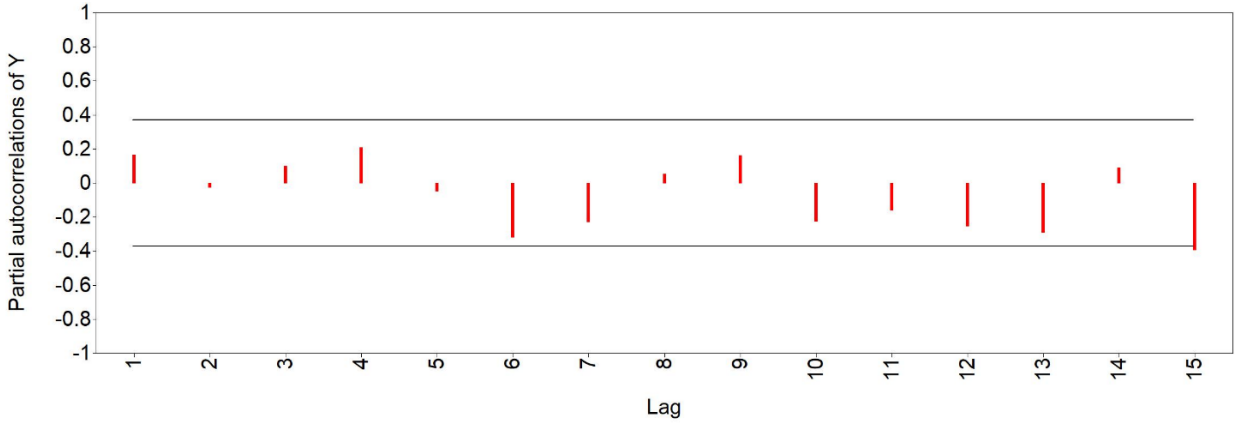
For DSEM models, Mplus offers limited global fit indices consisting primarily of the Bayesian based information theory indexed (DIC) for model comparisons. Given this, I use somewhat ad hoc methods to gain a sense of model appropriateness. I do so after I ensure through examination of the Potential Scale Reduction indices (PSRs) that model convergence has been achieved (see Chapter 8). The highest PSR observed at the final iteration was 1.00, which suggests adequate convergence.

One potential source of specification error is if the effects of the treatment vary by time, i.e., there is a treatment by time interaction. To evaluate graphically if a model without an interaction term is reasonable, I used the program called *DSEM plot* on the *Programs* tab of my website to plot the raw mean sleep quality in the intervention and control groups for each of the 28 days. Here is the resulting plot:



The control group is in blue and the intervention group is in red. The horizontal dashed straight line represents the overall mean across the groups. The uneven dashed line is a smoother. The vertical distance between the two smooths reflects the mean difference between the conditions and you can see that it is roughly the same across time. The within time variability is thought to reflect random events that sometimes push scores above the overall group mean and sometimes push scores below the overall group mean. However, the system is such that these "pushes" eventually revert back or regress to the overall mean, so the overall mean is conceptually important. Looking at the plot, you see the differences between the conditions becomes negligible around day 10. But this is usually viewed as random noise unless there is a strong theoretical reason to expect it. For the outcome for sleep quality, the assumption of no treatment by time interaction seems to "fit the data" and is what I expected to be the case a priori. This also was true when I generated the DSEM plots for each of the mediators.

A second source of possible misspecification is the posited lagged effects in the within-person model. Perhaps my model should have included a second order lag or a third order lag. After the syntax is executed, Mplus provides a **partial autocorrelation function plot** that provides perspective on this issue. You obtain the plot with the output window open by clicking in the main Mplus interface the menu item **Plot > View plots > Time series plots**. Then click the **View** button choose **partial autocorrelation function plot**, and choose the variable you want to focus on. In this case, I analyze the autoregressive effect for Y. Here is the result:

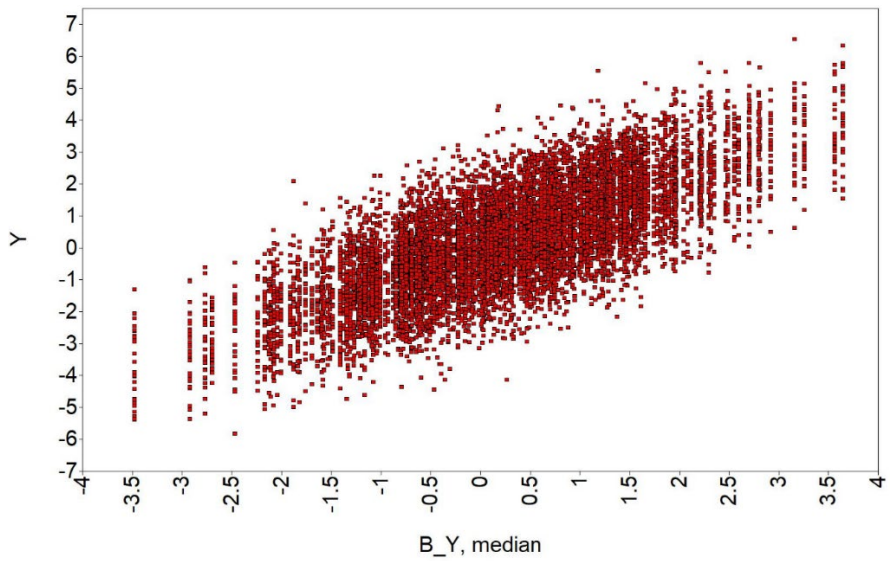


The two solid black lines at roughly ±0.40 represent 95% confidence intervals. Any red vertical bar that stays inside these lines is statistically indistinguishable from zero, i.e., the partial autocorrelation likely is random noise. Because the first-order autoregressive effect was explicitly accounted for in the DSEM within-person model, the

residual partial autocorrelation at lag 1 appropriately drops within the 95% confidence intervals. No systematic spikes emerged at higher lags, confirming that a first-order lag structure was sufficient. The lag 15 red line just barely crosses the 95% lower limit, but it makes no conceptual sense and is likely a chance result due to the number of tests performed. The model seems consistent with the data for the autoregressive lag for Y and this also was the case when I conducted a comparable analysis for M1 and M2.

As yet another specification-based test of model-data correspondence I can examine a scatterplot of the observed Y scores against the model's predicted scores. In a Bayesian DSEM framework, individual predicted values are derived from the posterior distribution of the model parameters. Rather than relying on a single deterministic point estimate, Mplus uses a MCMC sampling approach to generate a distribution of predicted values for each person-period observation. If the linear relationships, autoregressive carry-over effects, and between-person paths are correctly specified, a scatterplot of the residuals (the distance between observed and predicted median predicted scores) should display an even, random distribution around zero, without systematic curvature.

To apply this diagnostic in Mplus, you must request the generation of posterior draws by including `TYPE = PLOT3;` and `FACTORS = ALL (500);` within the `PLOT:` command block (see Line 34 of Table 16.19). This syntax tells Mplus to retain 500 samples from the posterior distribution for each latent and predicted component. After running the syntax, I navigate to the Mplus 'Plot' menu and select 'View Plot', followed by 'Scatterplots'. From the variable selection menu, choose the observed outcome Y for the vertical axis and the median of the posterior predicted values (labeled by Mplus as the median of B_Y) for the horizontal axis. Visual inspection of the resulting scatterplot can then be used to confirm that the residuals mirror random noise, signaling reasonable model-data correspondence. Here is the plot for the current data:



The patterning of residuals seems reasonable. Parenthetically, because the total number of observations is dominated by within-person data points (28 days times 400 participants = 11,200 compared to 400 between-person points), this scatterplot is primarily sensitive to misspecification in the within-person model dynamics. Substantive misspecifications in the between-person model often are obscured by the sheer volume of within-person data points.

A final check I can pursue is comparing (a) the model predicted correlations between variables at the within-person level with the observed correlations at the within-person level, and (b) the model predicted correlations between variables at the between-person level with the observed correlations at the between-person level. The model-implied correlations appear in the `Technical 4` section of the output. Here are the reported predicted within-person correlations:

```
       ESTIMATED WITHIN CORRELATION MATRIX FOR THE LATENT VARIABLES

              Y       M1      M2       Y&1     M1&1     M2&1

Y         1.000
M1        0.306    1.000
M2        0.289   -0.003    1.000
Y&1       0.347    0.095    0.089    1.000
M1&1      0.177    0.309   -0.001    0.306    1.000
M2&1      0.167   -0.001    0.309    0.289   -0.003    1.000
```

and here are the maximum likelihood based observed correlations:

```
      ESTIMATED WITHIN CORRELATION MATRIX FOR THE LATENT VARIABLES

              Y       M1      M2       Y&1     M1&1     M2&1

Y         1.000
M1        0.325    1.000
M2        0.315    0.010    1.000
Y&1       0.458    0.122    0.137    1.000
M1&1      0.261    0.415    0.037    0.324    1.000
M2&1      0.248    0.019    0.423    0.340    0.066    1.000
```

I describe how I calculated the observed correlations in Appendix B. Some methodologists argue that comparing the two matrices is like comparing apples to oranges. The model implied correlations are Bayesian based and adjusted for Ludtke's bias whereas the observed correlations are maximum likelihood based and derived from a different statistical theory. I don't disagree with this argument. However, when sample sizes are reasonable and the priors are uninformative, the Bayesian correlations usually converge with the maximum likelihood estimates. I look for blatantly large disparities that signal something is likely amiss. I do not use this as a fine-grained fit diagnostic. In the present case, lack of correspondence between correlations does not raise any red flags. This also was the case for the between-person correlations (see Appendix B).

Based on the above and coupled with the theoretical coherence of the results described below, I am comfortable moving forward with the analytic results.

Effect of the Intervention on the Outcome

Estimates of the overall effect of the intervention on sleep quality is based on treatment versus control differences for the average reported sleep quality across the 28 days. The FISEM estimate is obtained from the `MODEL CONSTRAINT` commands in Lines 31 to 35 of Table 16.19 with specific focus on `toteff` which is the sum of the indirect effects through the two mediators plus the direct effect of the treatment condition on sleep quality (see Line 35). Here is the relevant output:

```
Between Level

                          Posterior  One-Tailed          95% C.I.
               Estimate       S.D.     P-Value   Lower 2.5%  Upper 2.5%  Sig

New/Additional Parameters
INDM1             0.108      0.036       0.000       0.042       0.182     *
INDM2             0.180      0.047       0.000       0.093       0.274     *
TOTEFF            0.369      0.120       0.001       0.140       0.610     *
```

The implied mean difference between the intervention and control groups was 0.37 with a 95% credible interval of 0.14 to 0.61. Because the interval does not contain the value of 0, the effect is statistically significant, $p < 0.05$. Suppose that prior to the study, the program staff and research team set a meaningfulness standard of 0.25. Because the lower limit of the credible interval was less than this, we conclude that although the intervention effect was non-zero, we cannot confidently conclude the effect is meaningful. To be sure, the observed point estimate of 0.37 is suggestive of meaningfulness, but there is too much variability around the estimate in the posterior distribution to be fully confident that the program makes a meaningful difference.

I obtain the estimated means for the treatment and control groups from the `MODEL CONSTRAINT` commands in Lines 42-43 in Table 16.19. These commands were structured based on previous strategies in this chapter to define estimated means in a FISEM context. This involved using the intercepts and regression coefficients via the algebra of linear models to derive values for the estimated means. Here is the output for which I focus on `YMEANC` for the control group mean and `YMEANT` for the intervention mean:

		Posterior	One-Tailed	95% C.I.		
	Estimate	S.D.	P-Value	Lower 2.5%	Upper 2.5%	Sig
New/Additional Parameters						
INDM1	0.108	0.036	0.000	0.042	0.182	*
INDM2	0.180	0.047	0.000	0.093	0.274	*
TOTEFF	0.369	0.120	0.001	0.140	0.610	*
M1MEANC	0.067	0.072	0.173	-0.070	0.213	
M1MEANT	0.454	0.070	0.000	0.320	0.592	*
M2MEANC	0.062	0.080	0.220	-0.100	0.217	
M2MEANT	0.616	0.078	0.000	0.462	0.769	*
YMEANC	0.063	0.085	0.232	-0.104	0.230	
YMEANT	0.431	0.084	0.000	0.271	0.601	*

The model implied mean sleep quality in the intervention group was 0.43 ±0.17 and in the control group it was 0.06 ±0.17.

Effect of the Intervention on the Mediators

Here is the output for estimating the effect of the intervention on each mediator:

```
Between Level

                                Posterior  One-Tailed          95% C.I.
                  Estimate       S.D.     P-Value    Lower 2.5%  Upper 2.5%  Sig

M1          ON
 TREAT             0.387        0.100      0.000        0.187       0.583      *
 M1B               0.349        0.049      0.000        0.252       0.445      *
 COV1              0.369        0.047      0.000        0.278       0.464      *

M2          ON
 TREAT             0.554        0.112      0.000        0.330       0.769      *
 M2B               0.227        0.058      0.000        0.114       0.342      *
 COV1              0.265        0.053      0.000        0.160       0.368      *
```

The estimated mean difference between the intervention and control groups for M1 averaged across the 28 days was 0.39 with a 95% credible interval of 0.19 to 0.58. Because the interval does not contain the value of 0, the effect is statistically significant, $p < 0.05$. Suppose that prior to the study, the program staff and research team set a meaningfulness standard of 0.25. Because the lower limit of the credible interval was less than this, we conclude that although the intervention effect on the mediator was non-zero, we cannot confidently conclude the effect is meaningful in the population. As with the total effect, the observed point estimate of 0.39 is suggestive of meaningfulness, but there is too much variability around the estimate in the posterior distribution to be fully confident of a meaningful effect. The estimated mean M1 in the intervention group was 0.45 (95% credible interval = ±0.32 to 0.59) and in the control group it was 0.07 (95% credible interval = -0.07 to 0.21) per the prior table of means.

The mean difference between the intervention and control groups for M2 was 0.55 with a 95% credible interval of 0.33 to 0.77. Because the credible interval does not contain the value of 0, the effect is statistically significant, $p < 0.05$. Suppose that prior to the study, the program staff and research team set a meaningfulness standard of 0.25. Because the lower limit of the credible interval is larger than the meaningfulness standard, we conclude the effect is meaningful in the target population. The mean M2 in the intervention group was 0.62 (95% credible interval = ±0.46 to 0.77) and in the control group it was 0.06 (95% credible interval = -0.10 to 0.22; see `M1MEANT` and `M1MEANC` in the prior table of means)..

Effect of the Mediators on the Outcome

The estimated effect of the mediators on the outcome is evaluated at two levels, the within-person level and the between person-level. Here are the results for the within-person regression of sleep quality on the two mediators and the lagged sleep quality:

```
Within Level

                    Posterior  One-Tailed         95% C.I.
             Estimate   S.D.     P-Value   Lower 2.5%  Upper 2.5%  Sig

Y        ON
  M1          0.308    0.009      0.000       0.290       0.326    *
  M2          0.290    0.009      0.000       0.272       0.308    *
  Y&1         0.297    0.009      0.000       0.279       0.314    *
```

Holding constant carry-over effects of the prior night's sleep on the current night's, sleep, M1 yielded a path coefficient of 0.31 (95% CI = 0.29 to 0.33) for predicting sleep quality. This represents a statistically significant effect, $p < 0.05$. Suppose that prior to the study, the program staff and research team decided a reasonable meaningfulness standard was a population coefficient of 0.25. Because the credible interval lower limit for M1 exceeded this value, the presumed effect is deemed meaningful in the target population.

The same conclusions held for M2. Holding constant carry-over effects of the prior night's sleep on the current night's, sleep, M2 yielded a path coefficient of 0.29 (95% CI = 0.27 to 0.31) for predicting sleep quality. This represents a statistically significant effect, $p < 0.05$. Because the lower limit of the credible interval for M1 exceeds the meaningfulness standard, the effect is deemed meaningful in the target population.

Here is the output for evaluating the estimated effects of the mediators on sleep quality at the between-person level:

```
Between Level

                    Posterior  One-Tailed         95% C.I.
             Estimate   S.D.     P-Value   Lower 2.5%  Upper 2.5%  Sig

Y        ON
 M1          0.286    0.056      0.000       0.179       0.399    *
 M2          0.329    0.051      0.000       0.230       0.429    *
 TREAT       0.076    0.115      0.255      -0.150       0.303
 COV1        0.114    0.057      0.024       0.002       0.227    *
```

Note that the coefficients for M1 and M2 are comparable in magnitude to what they were for the within-person analyses. Both were statistically significant, $p < 0.05$. However, the credible intervals for the coefficients are much wider than those for the within-person

level, yielding a scenario where the lower limit of the 95% credible interval at the between-person level failed to exceed the meaningfulness standard for both M1 and M2. This likely is because the within-person coefficients are based on a much larger N and have less sampling error.

Taking into account both sets of analyses, I would include that both M1 and M2 are relevant to the sleep quality in non-trivial ways. Note also that the direct effect of the treatment condition on sleep quality independent of M1 and M2 was not statistically significant; the 95% credible interval is -0.15 to 0.30 and contains the value of zero.

Omnibus Mediation

As noted multiple times, my preference for purposes of program evaluation is to evaluate mediation relative to careful analysis of the separate links in each mediational chain as opposed to omnibus mediation tests. For those who prefer or who also want to pursue omnibus mediation tests, one possibility is to use the joint significance test (JST) described throughout this chapter. In the current case, both M1 and M2 satisfied the criteria for rejecting the null hypothesis of no mediation in that in both cases, the path coefficients for T→M and for M→Y were statistically significant. The traditional coefficient product tests derive from the `MODEL CONSTRAINT` commands in Lines 34 and 35 of Table 16.19. Here is the relevant output:

```
                    Posterior  One-Tailed          95% C.I.
             Estimate       S.D.     P-Value   Lower 2.5%  Upper 2.5%  Sig

New/Additional Parameters
INDM1           0.108      0.036       0.000       0.042       0.182      *
INDM2           0.180      0.047       0.000       0.093       0.274      *
```

The estimated omnibus indirect effect for M1 was 0.11 with a 95% credible interval of 0.04 to 0.18. The estimated omnibus indirect effect for M2 was 0.18 with a 95% credible interval of 0.09 to 0.27. In both cases, the omnibus tests were statistically significant because the value zero was not contained in their respective credible intervals.

Overall Study Conclusions

Based on the above analyses, I conclude that the intervention had an effect on sleep quality but that although the data are suggestive, and I can't confidently conclude the effect is meaningful in the population. The intervention affected both M1 and M2 but the effects only rose to the level of likely population meaningfulness for M2. There was evidence that both M1 and M2 affected sleep quality to about the same extent in the

between-person and within-person analyses but these effects only rose to statistically reliable meaningfulness in the target populations for the within-person analyses. The was not compelling evidence for an independent effect of the intervention on the outcome over and above the two mediators.

Additional Considerations

Random or Person-Varying Coefficients

In DSEM methodologists make distinctions between fixed or person-invariant coefficients and random or person-varying coefficients. Fixed coefficients represent an effect (e.g., a path coefficient) that takes on the same value across all individuals in the sample. It estimates a single, population-average parameter by assuming there are no meaningful individual differences in the parameter. In the sleep quality example, we found that the within-person coefficient for the effect of M1 on Y was 0.31. For a fixed coefficient model, the working assumption is that this effect of M1 on Y is the same for all individuals in the population. In sample data, it may vary from one individual to the next, but this variation is just random noise. The true value of this coefficient is the same for everybody.

For a random coefficient, the coefficient in question is thought to vary from one individual to the next. In the sleep quality example, the effect of M1 on Y might differ from one individual to the next. In DSEM random coefficient models, if you plotted the different coefficient values across all individuals in the population, it is assumed they will be normally distributed. Robust estimation methods can accommodated non-normally distributed coefficients. Mplus essentially calculates a mean and variance for the varying coefficients and reports them on the output. You also can explore determinants of coefficient variation.

The sleep quality example used all fixed coefficients. The programming and output for a random coefficient model is quite different so I cover that in a separate document called *Random Coefficients in Dynamic Structural Equation Modeling* on the *Resources* tab of my webpage under Chapter 16.

Number of Time Points

It generally is recommended that DSEM have a minimum of 10 repeated measures (Time > 9) to accurately separate within-person dynamics from between-person dynamics. Fewer time points frequently leads to estimation failure and parameter bias, especially for the case of models with random slopes. Some methodologists recommend a minimum of 20 repeated measures. As noted, Mplus relies on Markov Chain Monte Carlo (MCMC)

sampling. With fewer than 10 time points, the iterative estimation process frequently struggles to converge, or it produces unstable estimates of within-person parameters.

If you have fewer than 10 repeated measures, it probably is best to use the approaches described in the first numerical example in this chapter, which emphasized Allison's SEM-based fixed effects approach for the analysis of M→Y links. Allison's approach evolves out of econometrics and sociology and treats each individual's intercept as a *fixed parameter* in ways that mathematically sweep away between-person measured and unmeasured heterogeneity. DSEM, by contrast, treats the individual intercepts as *random parameters* that follow a probability distribution, usually assumed to be normally distributed in the population. It relies on latent centering and decomposition methods to remove *between-person* variance from the within-person estimates while still allowing for tests of between-person effects in a full information estimation sense. A thorough RET analysis that uses Allison's methods typically requires LISEM to address each of the three core questions surrounding mediation for an RET.

Residual Dynamic Structural Equation Modeling

When people are measured many times in close proximity the outcome (and mediators) often will be autocorrelated. Usually, measures taken closest together are most related with the strength of that relationship then decaying as the measures become further apart in time. In our example, we modeled the dependency by representing it as carry over effects via a first order lagged outcome or mediator. Mplus has recently introduced a specialized form of DSEM called **Residual DSEM** (RDSEM). It represents an alternative approach to dealing with dependencies between assessments in time series data. In standard DSEM, dependencies are accommodated by introducing causal lags for variables. RDSEM differs structurally from DSEM in the sense that autoregressive relationships operate on the *residuals* of the variable in question. Instead of modeling the outcome Y at time *t* as a function of the previous outcome Y at time *t-1*, RDSEM models Y at time *t* as a function of concurrent predictors (i.e., mediators) at time *t* and then accounts for the across-time dependencies in measures by allowing their disturbances to be autocorrelated or autoregressively related. DSEM models lag effects directly on the observed or latent variables themselves whereas RDSEM models lag effects strictly within the disturbance terms. This distinction changes the substantive meaning of your path coefficients and dictates a different set of mathematics underlying your model.

In terms of model equations, in DSEM the autoregressive structure is built directly into the primary structural equation:

$$Y_t = a + b1\ M_t + b2\ Y_{t-1} + d_t$$

with b1 representing the effect of M_t on Y_t holding Y_{t-1} constant and *d* is the usual disturbance term. This representation prevents the estimate of the contemporaneous path b1 from being muddied by the impact of the history of Y by conditioning on Y_{t-1}. In RDSEM, the structural equation remains completely contemporaneous as the time-series dependency is pushed into a secondary disturbance equation:

$$Y_t = a + b1\ M_t + d_t$$

$$d_t = b2\ d_{t-1} + e_t$$

where *e* is a disturbance term among the errors. In this case, b1 represents the pure, concurrent structural relationship between M_t and Y_t. The autoregressive parameter acts as an "auxiliary" filter to clean up the autocorrelation without changing the core meaning of b1.

The choice between using DSEM or RDSEM largely depends on your theoretical question and the nature of your data. Researchers use DSEM when time-lagged causal processes, auto-regressive or otherwise, are of substantive interest, such as wanting to know if an increase in morning anxiety increases later afternoon anxiety or later afternoon fatigue. RDSEM is more likely to be used when evaluating concurrent, time-specific effects, such as the immediate impact of a change in a mediator on a change in the outcome. It is well suited to handling concurrent, same-day causal relations without confounding them with previous time points but still accounting for dependencies in the scores across time. The coefficient b1 in DSEM is a conditional effect, namely conditional on Y_{t-1}. The coefficient in RDSEM is more akin to a marginal effect averaged across the past state(s).

Although I do not treat RDSEM in this book, it is well worth mastering. I provide references in the next section and on the *Resources* tab of my webpage in Chapter 16.

Concluding Comments on DSEM

Dynamic structural equation modeling is a powerful tool for analyzing intensive longitudinal data. It provides perspectives from both a within-person perspective and a between-person perspective, both of which can be useful. One of the most commonly applied analytic methods to longitudinal RCTs has been what are commonly called mixed model analyses (Singer & Willet, 2003). DSEM and RDSEM in Mplus are far more powerful and flexible than traditional linear mixed-effects models as implemented in R packages like *nlme* and *lme4*. While traditional linear mixed models for intensive data analyses are limited to univariate models, DSEM and RDSEM integrate time-series

analysis, multilevel modeling, and structural equation modeling (SEM) to offer distinct advantages, including mediational modeling and latent variable modeling, among others.

I have provided only a simplified introduction to DSEM. I encourage you to explore it in more depth. For useful tutorials and references, see Asparouhov and Muthén (2020a, b; 2026); Asparouhov, Hamaker and Muthén (2018); Hamaker and Asparouhov (2019); Langenberg, Helm, McCabe, Günther and Mayer (2026); McNeish and MacKinnon (2025); McNeish, Mackinnon, Marsch & Poldrack (2021); Muthén, Asparouhov & Keijsers (2025); and Wang and Kim (2024).

For interesting applications of DSEM to single-person designs (N=1), see Miočević et. al., (2022). For a thoughtful paper on comparing within-person effects with between person effects, see Neubauer, Koval, Zyphur & Hamaker (2025). For discussions of applying DSEM to binary outcomes, see McNeish, Somers and Savord (2024) and Berli, Inauen, Stadler, Scholz and Shrout (2021). For extensions of the method to scale-location models that develop prediction models of variances rather than means in intensive longitudinal data, see Hedeker, Mermelstein, & Demirtas, (2008).; Williams, Martin, Liu & Rast (2020); and Zhang & Hedeker (2023). For applications where the intervention occurs repeatedly over time (such as being provided an app to be accessed daily on one's cell phone, with accessing the app on a given day treated as intervention administration), see McNeish and MacKinnon (2025).

LATENT CHANGE SCORE MODELS

Another approach to analyzing changes over time in an RET is the **latent change score model** (LCSM; McArdle, 2009; Klopack & Wickrama, 2020; Hilley. & O'Rourke, 2022). In a LCSM, latent change is parameterized as within-person changes in the outcome but the means and variances of the scores function as between-person parameters. This is analogous to the latent growth curve dynamics discussed above. LCSMs have been used in a wide variety of contexts. My focus here is on how they can be applied to RETs. The approach has been used to analyze RCTs in the published literature but usually in limited ways with relatively few time periods (e.g., McArdle & Prindle. 2008; Serang, 2026; Yi & Serang, 2026). RCT studies that have focused on mediation using the LCSM have generally used simple pre-post designs (e.g., Valente & MacKinnon, 2017, 2021; Fishbein et al., 2022; Helland et al. 2023), but there are exceptions (e.g.,, Goldsmith et al., 2018).

LCSM prioritizes the concept of change scores. The core logic derives from simple algebraic manipulation of a change score. The change from the variable Yt to Y$t+1$ can

be defined as (Y*t+1* - Y*t*). This expression can be reframed to predict Y*t+1* via algebraic manipulation as:

Y*t+1* = Y*t* + (Y*t+1* - Y*t*)

which I also can express as

Y*t+1* = Y*t* + tau

where tau is the operative residual Y*t+1* - Y*t* in the prior equation. I can add regression coefficients to the right hand side of the above equation with the constraint that the coefficients must equal 1:

Y*t+1* = b_1 Y*t* + b_2 tau given $b_1 = b_2 = 1$

In words, this equation states: *"To predict* Y*t+1, take the starting score* Y*t multiplied by 1 and add the change score tau multiplied by 1 to it."* I can express this dynamic in an influence diagram as follows:

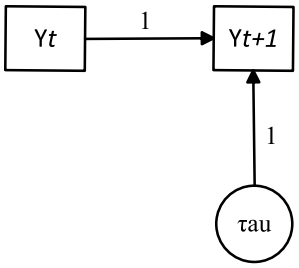


FIGURE 16.26 LCSM representation of change

In LCSMs, instead of working with calculated change scores per se, the above representation is used for modeling purposes and theory tests focused on change (tau) at different time points. To give this intuitive meaning using an everyday analogy to bank accounts,

Y*t* is your baseline bank account balance $100

Y*t+1* is your new bank account balance $150

tau is the *net change* to the account balance (a deposit of $50)

Tau represents the "deposit" or "withdrawal" that caused the baseline score to

change. Rather than comparing the first two set of numbers blindly, the interest instead is on the deposit. Such is the case in a LCSM. Essentially, tau becomes an outcome or a predictor and is captured by the logic of Figure 16.26. We do not work with calculated change scores. Rather, we work with a latent variable, tau, that captures change.

In SEM, tau is called a **phantom latent variable** that is used to isolate and represent the amount of change that occurs between two time points. Tau acts as an algebraic placeholder that forces the model to calculate a difference score while keeping the model's structural fit intact. In the vocabulary of Mplus, Y*t+1* is regressed onto Y*t* (with an autoregressive path fixed to 1.0) while also serving as an "indicator" of the phantom latent variable tau whose path to Y*t+1* is fixed to 1.0, per Figure 16.26.[13] Tau is considered to be a phantom latent variable because it has no indicators of its own and does not add new data to the model. It simply reparametrizes the existing relationship between Y at time 1 and time 2 to allow one to directly predict the resultant *growth* or *decline* using other predictors or to use the growth or decline to predict future outcomes.

A common use of LCSMs is to document change trajectories for longitudinal data much like latent growth curve models described earlier but it uses different parameterizations. Figure 16.27 illustrates an LCSM for a five wave longitudinal design.

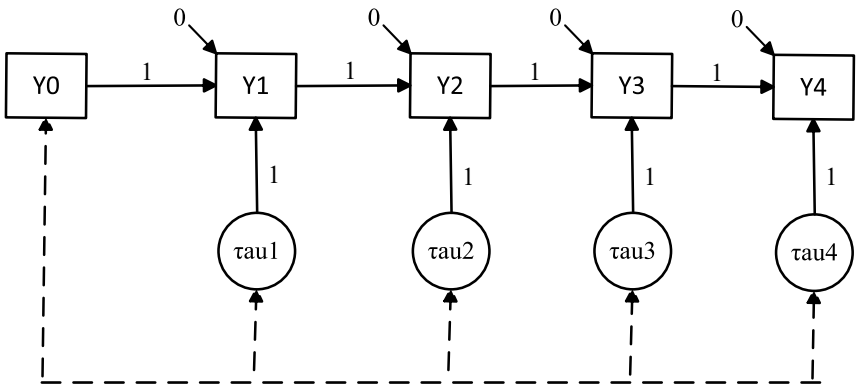


FIGURE 16.27 LCSM for five longitudinal time points F

For the correlations among the exogenous variables, the dashed line is a short hand way of signifying that all of them are to be correlated with each other; rather than drawing all the traditional pairwise curved arrows, I economize the diagram to reduce clutter. What is typically of interest in the above longitudinal model are the mean values

[13] The error score for Yt+1 is fixed to zero for identification purposes as is the autoregressive intercept from Yt to Yt+1 and the intercept from tau to Yt+1.

of the various tau. Each one indicates by how much the mean value of Y at time *t* shifts upward or downward from the previous mean of Y at time *t-1*. Like LGC models, we are analyzing trajectories but we do so by breaking the trajectories into sequenced steps.

I can add a time-varying predictor to this model, that I will call M, to determine if the stepped changes in M contemporaneously predict the stepped changes in Y. Figure 16.28 shows the essence of such a model, omitting disturbance terms and correlations to reduce clutter but which are estimated in ways I describe below.

There are many different forms that models of this type can take. I illustrate here the application of LCSMs to an RET in the spirit of Figure 16.28 but using a strategy known as **multi-group SEM**. This strategy simultaneously but separately fits the model in the intervention group and the control group and then formally contrasts model parameters in the two groups to answer our standard three questions for mediational modeling in RETs.

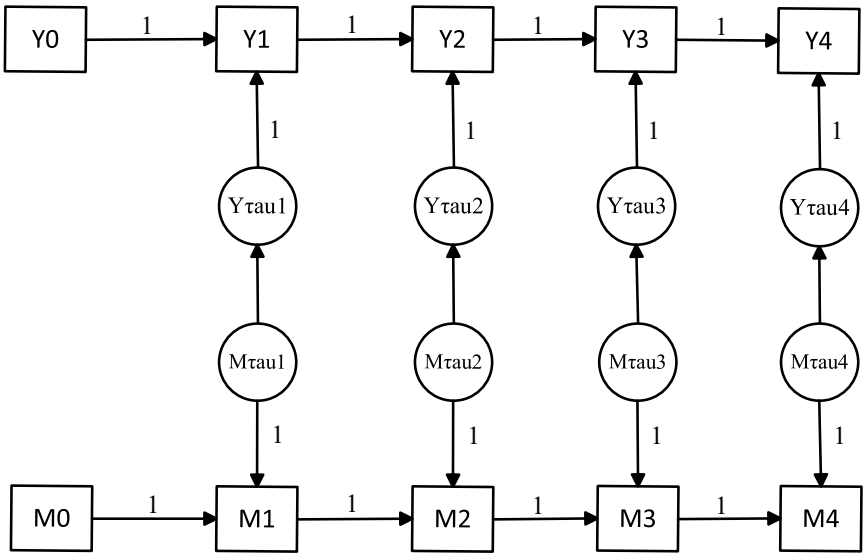


FIGURE 16.28 LCSM model with time varying predictor

Multiple Group Modeling

I adopt the multi-group strategy for LCSMs because it offers more flexibility than a single group SEM, it allows me to deal more easily with complex parameterizations, and it is simpler for introducing the `MODEL CONSTRAINT` commands in Mplus that I eventually invoke. The primary disadvantage of this approach is that it is somewhat more sample size demanding than single group SEM. Muthén and Curran (1997) advocate for the multi-group approach for RCTs in longitudinal contexts and recommend a sequence

of steps to use when pursuing it, such as conducting preliminary analyses on the intervention and control groups separately and testing for treatment-mediator interactions via across-group contrasts. My focus here is on explicating the basic set-up of the multigroup approach. See Muthén and Curran (1997) for elaborations of their recommendations.

I discuss SEM-based multiple group modeling in general and for purposes of moderator analysis in Chapter XX. To illustrate the approach to LCSMs for mediation analysis, I use Figure 16.28 as the organizing framework. In the analysis, I split the sample in two, half consisting of everyone in the intervention condition and the other half consisting of individuals in the control condition. I then apply the model in Figure 16.28 to each half. By doing so, the treatment condition is held constant in each group rendering it irrelevant to the within-group analyses. To be sure, I need to conduct comparative analyses across the intervention and control groups to answer the three core mediation-based questions of RETs. I show how to do this shortly. The baseline Y and M in the figure are represented by Y0 and M0 in each group. The post intervention variables (Y1 to Y4) and (M1 to M4) represent the immediate posttest and 3 follow-up periods each separated by 3 months.[14] For this hypothetical study, both M and Y are scored from 0 to 100 on 100 point metrics. M is the motivation to engage in a healthy lifestyle and Y is future healthy behavior performance. To make the scores somewhat easier to work with, I divided them by 10 so that the analytic metric is from 0 to 10.

Instead of physically conducting separate analyses on the two halves of the sample, SEM offers an option to combine the groups into a single analysis while still respecting the group differences, hence the name **multi-group analysis**. Mplus reports the results for the parameter estimates and a subset of fit indices for each group separately, but it also presents fit statistics for the two groups considered simultaneously. For example the overall chi square statistic for the simultaneous analysis equals the value of the chi square statistic for the first group plus the chi square statistic for the second group. The degrees of freedom for the overall chi square statistic equals the degrees of freedom associated with the chi square statistic for the first group plus the degrees of freedom for the chi square statistic for the second group. Simultaneous fit indices are also calculated for most of the other commonly reported fit indices, as I show below.

Table 16.20 presents the first portion of the Mplus syntax I used to execute the analysis. I add more commands using the `MODEL CONSTRAINT` option shortly. I omit them here in order to obtain more detailed global fit information because Mplus does not allow you to request modification indices when using the `MODEL CONSTRAINT` command.

[14] The time intervals do not need to be equally spaced in LCSMs unless theory dictates otherwise.

Table 16.20. Mplus Syntax for LCSM Example

```
1.  Multi-group LCSM Analysis;
2.  DATA:  FILE = lcsmm.dat;
3. VARIABLE:
4.  NAMES = y0 y1 y2 y3 y4 m0 m1 m2 m3 m4 treat;
5.  USEVARIABLES = y0 y1 y2 y3 y4 m0 m1 m2 m3 m4;
6.  GROUPING = treat (0 = control 1 = intervention);
7. ANALYSIS:
8.  ESTIMATOR = MLR;
9. MODEL:
10.  !DEFINE LATENT TRUE SCORES & FIX INTERCEPTS
11.  y1 ON y0@1; y2 ON y1@1; y3 ON y2@1; y4 ON y3@1;
12.  m1 ON m0@1; m2 ON m1@1; m3 ON m2@1; m4 ON m3@1;
13.  dy1 BY y1@1; dy2 BY y2@1; dy3 BY y3@1; dy4 BY y4@1;
14.  dm1 BY m1@1; dm2 BY m2@1; dm3 BY m3@1; dm4 BY m4@1;
15.  y1-y4@0; m1-m4@0;
16.  [y1-y4@0]; [m1-m4@0];
17.  !OVERRIDE AUTOMATIC COVARIANCES & PERMIT SPECIFIC ONES
18.  dy1-dy4 WITH dy1-dy4@0;
19.  dm1-dm4 WITH dm1-dm4@0;
20.  dy1-dy4 WITH dm1-dm4@0;
21.  !RE-SPECIFY VARIANCES AND COVARIANCES
22.  y0 WITH m0;
23.  y0 WITH dy1 dy2 dy3 dy4;
24.  m0 WITH dm1 dm2 dm3 dm4;
25.  y0 WITH dm1; m0 WITH dy1;
26.  y0 m0; [y0 m0];
27.  dy1-dy4; dm1-dm4;
28.  dy1 WITH dy2*; dy2 WITH dy3*; dy3 WITH dy4*;
29.  dm1 WITH dm2*; dm2 WITH dm3*; dm3 WITH dm4*;
30.  !SPECIFY THE CORE REGRESSIONS
31.  dy1 ON dm1 ; dy2 ON dm2 ; dy3 ON dm3 ;  dy4 ON dm4 ;
```

Most of the syntax should be familiar. On Line 6, I added syntax to identify for Mplus the variable used to define the two groups and in parentheses, what the numbers are on that variable that define the groups (in this case, 0 and 1) concurrent with labels to assign to the numbers (in this case "control" and "treat"). I exclude `TREAT` from the `USEVARIABLES` line because technically it is not used in the model; it only is used to define the two groups.

In multiple group SEM, it is understood that all of the lines under the word `MODEL` (Lines 9 to 31) apply to all of the groups unless you specifically tell Mplus to override some of them in some of the groups. For example, if you want to assign unique labels to a subset of the parameters in each group, you need to tell Mplus that is the case. Or, if you want to fix a parameter to a specific value in one group but not the other, you need to

inform Mplus of this. The way we communicate group specific qualifications to Mplus is to add to the above a separate `MODEL` command for each group containing the commands we want to take precedence over the more general `MODEL` commands. Before showing you this feature, I want to explain more of the above syntax.

Lines 10 to 16 define the latent taus in accord with Figure 16.26. Lines 11 and 12 specify the autoregressions of Y*t* onto Y*t-1* with fixed coefficients of 1 and M*t* onto M*t-1* with fixed coefficients of 1. Lines 13-14 regress Y*t-1* onto the pseudo latent variable taus which I label as `dy1` through `dy4` and `dm1` through `dm4`. Line 15 fixes the error variances of the observed Y and M to zero and the measurement thresholds to 0 (per footnote 13).

The Mplus default is to estimate the covariances between the latent variables which can lead to under-identified models for LCSMs with mediators. Lines 17 to 20 override these defaults by fixing the covariances to zeros. In subsequent statements, I use theory to help me choose which of these covariances to estimate and the constraints to impose on them. This occurs in the group specific syntax I show you below.

Lines 22 to 25 correlate the baseline exogenous variables with each other and with their respective latent change variables. Line 26 estimates the means and variances of the baseline variables. Line 27 estimates the variances of the pseudo latent variables.

Lines 28 and 29 are a step towards addressing my deactivation of Mplus defaults that forced the covariances among the M phantom variables and the disturbances for the Y phantom variables to be zero. Theoretically, I want to allow for correlations between some but not all of these variables. My goal is to be more parsimonious and theory driven about their estimation while also avoiding possible under-identification. Consider the correlations among the `dy1` to `dy4` disturbance variances. Fixing some elements of a covariance matrix to zero while leaving others free is known as **banding** or **lag-specific constraints.** One instantiation of banding is called **banded Toeplitz**. This strategy constrains covariances such that all adjacent waves share the same value (Cov(1,2) = Cov(2,3) = Cov(3,4) with non-adjacent covariances set to zero. Another approach uses a **heterogenous banded structure**. A **lag-2 heterogenous autoregressive structure** allows the adjacent covariances to differ from one another but again forces non-adjacent covariances to be zero. I used a variant of this approach. Based on theory and preliminary analyses, I expected the correlations between the adjacent disturbances to be weak, around 0.15, and quite weak (near zero) for non-adjacent categories. Rather than fixing non-adjacent covariances to zero, I decided to estimate them to allow for some association. However, to minimize non-identification I forced them to be equal. If this implementation is misguided and has non-trivial modeling consequences, the result should be poor model fit indices or it should reveal itself in sensitivity analyses. Lines 27 and 28 allow the correlations for the adjacent waves within each group to be estimated. I

add the * after each parameter to ensure Mplus estimates it by overriding my prior statements to fix them at zero. This is necessary. I introduce the equality constraints for the non-adjacent waves in the group specific models below. Line 31 specifies the core regression equations for the latent phantom variables.

Here is the additional syntax for each group where I supplement, override, or add group-specific labels to the control and treatment groups, respectively. In each of the MODEL commands, I identify the group using the labels I provided in Line 7:

```
32. MODEL control:
33.   dy1 ON dm1 (by1_c);
34.   dy2 ON dm2 (by2_c);
35.   dy3 ON dm3 (by3_c);
36.   dy4 ON dm4 (by4_c);
37.   [dy1] (ay1_c); [dy2] (ay2_c); [dy3] (ay3_c); [dy4] (ay4_c);
38.   [dm1] (mm1_c); [dm2] (mm2_c); [dm3] (mm3_c); [dm4] (mm4_c);
39.   dy1 with dy3* (cycov); dy1 with dy4* (cycov); dy2 with dy4* (cycov);
40.   dm1 with dm3* (cmcov); dm1 with dm4* (cmcov); dm2 with dm4* (cmcov);
41. MODEL intervention:
42.   dy1 ON dm1 (by1_i);
43.   dy2 ON dm2 (by2_i);
44.   dy3 ON dm3 (by3_i);
45.   dy4 ON dm4 (by4_i);
46.   [dy1] (ay1_i); [dy2] (ay2_i); [dy3] (ay3_i); [dy4] (ay4_i);
47.   [dm1] (mm1_i); [dm2] (mm2_i); [dm3] (mm3_i); [dm4] (mm4_i);
48.   dy1 with dy3* (iycov); dy1 with dy4* (iycov); dy2 with dy4* (iycov) ;
49.   dm1 with dm3* (imcov); dm1 with dm4* (imcov);  dm2 with dm4* (imcov);
      OUTPUT: Samp StdYX MOD(4) Residual Tech4  ;
```

Lines 33 to 36 regress the outcome latent change scores onto the latent mediator scores and assign unique labels for the control group for each coefficient. I make use of these labels later in the MODEL CONSTRAINT commands. Lines 37 and 38 estimate the means/intercepts of the latent change scores and Lines 39 and 40 specify estimation of the non-adjacent correlations with the equality constraints operationalized via the use of common labels. A parallel set of commands occur for the intervention group.

I execute the above syntax so that I can evaluate the overall fit of the model. Once I am convinced I have a reasonable fitting model, I add MODEL CONSTRAINT commands to explore substantively interesting contrasts. I then execute the augmented syntax again but with changes to the OUTPUT line to eliminate the request for modification indices. This is because if I add MODEL CONSTRAINT commands, Mplus forces me to remove the request for modification indices and I do not want to sacrifice them in my initial analysis when evaluating model fit. Here are the MODEL CONSTRAINT commands I add just before the OUTPUT line in the prior syntax for the second analysis:

```
MODEL CONSTRAINT:
  NEW(dy1_c dy2_c dy3_c dy4_c dy1_i dy2_i dy3_i dy4_i
  dm1_c dm2_c dm3_c dm4_c dm1_i dm2_i dm3_i dm4_i
  g_diff1 g_diff2 g_diff3 g_diff4 tc1v2 tc1v3 tc1v4 tc2v3 tc2v4
  tc3v4 ti1v2 ti1v3 ti1v4 ti2v3 ti2v4 ti3v4
  ey0_c ey1_c ey2_c ey3_c ey4_c ey0_i ey1_i ey2_i ey3_i ey4_i
  em0_c em1_c em2_c em3_c em4_c em0_i em1_i em2_i em3_i em4_i
  ey10_i ey20_i ey30_i ey40_i ey10_c ey20_c ey30_c ey40_c
  diffey10 diffey20 diffey30 diffey40
  em10_i em20_i em30_i em40_i em10_c em20_c em30_c em40_c
  diffem10 diffem20 diffem30 diffem40 );
!MEAN PHANTOM Y VARIABLES FOR CONTROL AND INTERVENTION GROUPS
  dy1_c = ay1_c + (by1_c * mm1_c);
  dy2_c = ay2_c + (by2_c * mm2_c);
  dy3_c = ay3_c + (by3_c * mm3_c);
  dy4_c = ay4_c + (by4_c * mm4_c);
  dy1_i = ay1_i + (by1_i * mm1_i);
  dy2_i = ay2_i + (by2_i * mm2_i);
  dy3_i = ay3_i + (by3_i * mm3_i);
  dy4_i = ay4_i + (by4_i * mm4_i);
!MEAN PHANTOM M VARIABLES FOR CONTROL AND INTERVENTION
  dm1_i = mm1_i;  dm2_i = mm2_i;  dm3_i = mm3_i;  dm4_i = mm4_i;
  dm1_c = mm1_c;  dm2_c = mm2_c;  dm3_c = mm3_c;  dm4_c = mm4_c;
!GROUP DIFFERENCES IN EFFECT OF M ON Y AT EACH WAVE
  g_diff1 = by1_i - by1_c;   g_diff2 = by2_i - by2_c;
  g_diff3 = by3_i - by3_c;   g_diff4 = by4_i - by4_c;
!CONTROL GROUP DIFFEENCES IN PATHS ACROSS TIME
  tc1v2 = by1_c - by2_c;    tc1v3 = by1_c - by3_c;
  tc1v4 = by1_c - by4_c;    tc2v3 = by2_c - by3_c;
  tc2v4 = by2_c - by4_c;    tc3v4 = by3_c - by4_c;
!INTERVENTION GROUP DIFFERENCES IN PATHS ACROSS TIME
  ti1v2 = by1_i - by2_i;    ti1v3 = by1_i - by3_i;
  ti1v4 = by1_i - by4_i;    ti2v3 = by2_i - by3_i;
  ti2v4 = by2_i - by4_i;    ti3v4 = by3_i - by4_i;
!MEANS FOR CONTROL GROUP
  ey0_c = int_y0_c; ey1_c = ey0_c + dy1_c;
 ey2_c = ey1_c + dy2_c; ey3_c = ey2_c + dy3_c;
 ey4_c = ey3_c + dy4_c; em0_c = int_m0_c;
 em1_c = em0_c + dm1_c; em2_c = em1_c + dm2_c;
 em3_c = em2_c + dm3_c; em4_c = em2_c + dm4_c;
!MEANS FOR INTERVENTION GROUP
 ey0_i = int_y0_i; ey1_i = ey0_i + dy1_i;
 ey2_i = ey1_i + dy2_i; ey3_i = ey2_i + dy3_i;
 ey4_i = ey3_i + dy4_i; em0_i = int_m0_i;
 em1_i = em0_i + dm1_i; em2_i = em1_i + dm2_i;
 em3_i = em2_i + dm3_i; em4_i = em3_i + dm4_i;
!DIFFERENCES FROM BASELINE FOR OUTCOME
 ey10_i = ey1_i - ey0_i ;  ey20_i = ey2_i - ey0_i ;
 ey30_i = ey3_i - ey0_i ;  ey40_i = ey4_i - ey0_i ;
```

```
99.  ey10_c = ey1_c - ey0_c ;  ey20_c = ey2_c - ey0_c ;
100. ey30_c = ey3_c - ey0_c ;  ey40_c = ey4_c - ey0_c ;
101. diffey10 = ey10_i - ey10_c ;  diffey20 = ey20_i - ey20_c ;
102. diffey30 = ey30_i - ey30_c ;  diffey40 = ey40_i - ey40_c ;
103. !DIFFERENCES FROM BASELINE FOR OUTCOME
104. em10_i = em1_i - em0_i ;  em20_i = em2_i - em0_i ;
105. em30_i = em3_i - em0_i ;  em40_i = em4_i - em0_i ;
106. em10_c = em1_c - em0_c ;  em20_c = em2_c - em0_c ;
107. em30_c = em3_c - em0_c ;  em40_c = em4_c - em0_c ;
108. diffem10 = em10_i - em10_c ;  diffem20 = em20_i - em20_c ;
109. diffem30 = em30_i - em30_c ;  diffem40 = em40_i - em40_c ;
```

Lines 61-69 estimate the means of the latent Y phantom variables at each time point (starting with the immediate posttest) for the intervention and control groups separately. Labels ending in `c` are for the control group and those ending in `i` are for the intervention group. Many of the statements in the `MODEL CONSTRAINTS` section use labels assigned to parameters in the primary syntax. The logic throughout this section follows directly from the logic I have outlined for prior topics in this chapter and in the document *Additional Applications of Latent Growth Curve Modeling* on the *Resources* tab of my webpage. Lines 70 to 72 do the same for the mediator. Lines 73 to 75 compare the effect of changes in M on changes in Y for the treatment versus control groups to determine if there is a treatment by mediator interaction. Lines 76 to 79 test if the effect of changes in M on changes in Y vary by wave for the control group. Lines 80 to 83 do so for the intervention group. Lines 84 to 95 calculate the model implied means of the observed Y and M at each wave. Lines 96 to 102 estimate mean differences between the outcome at a given post-intervention time period and the baseline outcome. It also tests the differences between the these mean improvement rates for the intervention versus control groups. Lines 103 to 109 performs the same operations but for the mediator. I show below how I make use of the different contrasts. There are other contrasts you might find of interest but the above are the ones I focus on here.

Model Fit

Here are the results for overall model fit based on the initial analysis:

```
MODEL FIT INFORMATION

     Chi-Square Test of Model Fit

               Value                             40.114*
               Degrees of Freedom                    38
               P-Value                           0.3766
```

```
Chi-Square Contribution From Each Group

          CONTROL                            25.845
          INTERVENTION                       14.269

RMSEA (Root Mean Square Error Of Approximation)

          Estimate                            0.011
          90 Percent C.I.                     0.000  0.034
          Probability RMSEA <= .05            1.000

CFI/TLI

          CFI                                 1.000
          TLI                                 0.999

SRMR (Standardized Root Mean Square Residual)

          Value                               0.019
```

The overall chi square index is for the two groups combined. The p value was statistically non-significant suggesting a reasonable fitting model. As noted, the total chi square statistic is an additive function of the separate chi square indices for each group. In the section `Chi-Square Contribution From Each Group`, Mplus reports the separate group chi square values. One typically does not want to see a sizeable disparity in these values because it then suggests the model is fitting much better in one of the groups. If the disparity is large, the question becomes why is the model fitting worse in one group than the other and you will want to resolve this. In this case, the disparity, coupled with information garnered from other diagnostics, is not unreasonable.[15]

The remaining indices should all be familiar to you but keep in mind that these statistics are for the simultaneous analysis, not the per group analyses. All of them appear to be in order for the current example. For the localized fit indices, Mplus reports modification indices, cell-by-cell z tests of the difference between the predicted and observed covariances, as well as estimates of the differences between the predicted and observed correlations *per se*. I do not show these in the interest of space, but they follow the same format and interpretation as prior chapters. In the current case, there were no theoretically meaningful problematic localized fit indices for either of the groups.

In general, if you make a modification in one of the groups to free-up estimation of a given parameter that was previously constrained to zero in the model (e.g., introduce a

[15] You can obtain separate indices for each group of the RMSEA, CFI, and SRMR indices but to do so you must execute separate analyses on each group via a traditional Mplus program. I did so in this case and the fits in both groups were reasonable.

correlated disturbance that had before been set to zero), you will need to also free up that particular parameter in the other group even if the local fit indices for the other group suggest the parameter can be left alone. In most multiple group analyses, we want the full model to be parallel in all the groups for the analytics to work properly. This is not an absolute requirement but it is common in practice. To allow different model configurations in the multiple groups requires specialized programming and non-trivial statistical issues that are beyond the scope of this. If a parameter is freed in a group where it did not need to be but is freed for purposes of creating parallel configurations, then all that is lost is a degree of freedom in that group that otherwise could have been spared. The value of the estimate itself when the modified program is executed likely will be a value near zero, which is what you assumed anyway when you omitted the parameter.

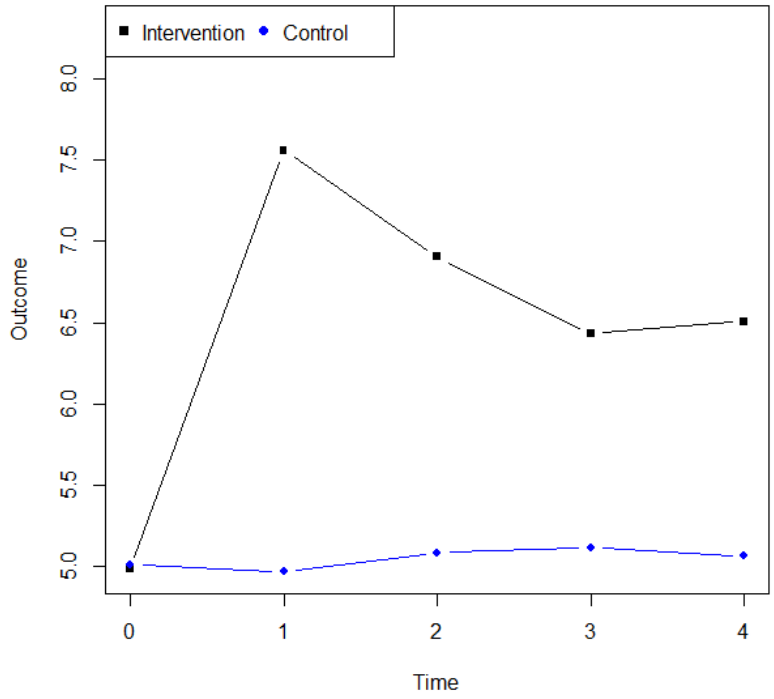
Given a reasonable fitting multi-group model, I then remove the request for modification indices from the `OUTPUT` line and add my `MODEL CONSTRAINT` commands. Before I review the results of the analysis, there is an important feature of multi-group SEM that I should mention. In traditional single group RET models, we obtain a single estimate of each model parameter. In multigroup-analyses we obtain two parallel parameter estimates, one for the intervention group and the other for the control group. When you conduct the single-group analysis, you implicitly make the assumption that the coefficients in the two groups are equal and that any difference between them when analyzed separately reflects random noise. The multi-group solution estimates them separately such that you can empirically evaluate this homogeneity assumption. I consider such tests below. Although the multi-group analysis has the ability to test the underlying homogeneity assumption, it comes at some cost because the separate parameter estimates for each group are based on a smaller N than the single group analysis, namely they each are based on the separate group sizes. For studies with small N in one of the groups, the multiple group strategy may not be viable.

A related difference between the single and multigroup approaches is that the latter allows *all* of the coefficients in the model for the two groups to vary between the groups, including covariates and the like. Some would argue that this property is a strength as the multi-group analysis is more nuanced given this property. Others argue the property can lead to sample-specific overfitting. Of course, you can always impose across-group equality constraints on whatever coefficients you wish but this should be done cautiously. Strict coefficient equality is a stringent condition that rarely holds in the real-world. Perhaps it is better, the argument goes, to allow for minor between-group deviations to more accurately reflect a world in which minor differences likely exist. If I have strong theory that suggests across group equivalence for a given coefficient, then I let that theory guide my equality constraint decisions.

With that, let's answer the three core questions of mediation analyses in RETs.

Total Effect of the Treatment on the Outcome

To evaluate the effects of the intervention on the outcome, I first examine treatment condition mean differences for Y at the immediate posttest. To help put this analysis and other contrasts in context, I find it useful to plot the predicted Y means for the two groups to gain an intuitive sense of the broader total effect dynamics. I use the program on my website called *Temporal line plot* to graph the model-implied means from the `MODEL CONSTRAINTS` output generated by Lines 99-100 and 105-106. Here is the plot:



Time 0 is the baseline and Time 1 is the immediate posttest. It is evident from the plot that the pre-post improvement in Y for the intervention group was larger than for the control group. Here is the output that documents the pre-to-post improvement rates as reflected in the phantom variables `DY1_I` and `DY1_C`:

```
                                                    Two-Tailed
                    Estimate       S.E.  Est./S.E.    P-Value

New/Additional Parameters

    DY1_I              2.578      0.060     43.047      0.000
    DY1_C             -0.041      0.059     -0.693      0.489
```

The estimated average pre-to-post improvement in healthy behaviors for the intervention group was 2.58 ±0.12 (CR = 43.05, $p < 0.05$) or about 26 points on the original 0 to 100 metric. For the control group, the average pre-to-post improvement was -0.04 ±0.12 (CR = 0.69, *ns*). I provide a formal test of the group differences below.

In longitudinal designs with multiple follow-ups there is more than one "total effect." In addition to asking if there are treatment-control differences in Y improvements from baseline to the immediate posttest, I also can ask the same question for each follow-up. I addressed this issue in the `MODEL CONSTRAINTS` commands by estimating the change from baseline for each follow-up period via Lines 96 to 102. Here are the results for the intervention and control groups (note: in the labels, `EY` stands for the expected value of Y, the first digit references the post-intervention time period, the second digit is 0 to reflect it is compared against the baseline and the intervention group is represented by `I` and the control group by `C`; `EY10_I` is the model-implied mean improvement from baseline to the first post-intervention for the intervention group; `EY20_I` is the implied mean improvement from baseline to the second post-intervention follow-up for the intervention group; and so on):

```
MODEL RESULTS

                                                    Two-Tailed
                  Estimate       S.E.  Est./S.E.    P-Value

New/Additional Parameters

    EY10_I           2.578      0.060     43.047      0.000
    EY20_I           1.922      0.075     25.768      0.000
    EY30_I           1.453      0.089     16.267      0.000
    EY40_I           1.525      0.099     15.365      0.000

    EY10_C          -0.041      0.059     -0.693      0.489
    EY20_C           0.074      0.074      0.993      0.321
    EY30_C           0.107      0.092      1.165      0.244
    EY40_C           0.052      0.106      0.491      0.623
```

The significance test for each row evaluates if the degree of improvement from baseline to the designated posttest/follow-up is different from zero. For example, the improvement from baseline to wave 4 for the intervention group was 1.52 ±0.20 which was statistically significantly different from zero (CR = 15.37, $p < 0.05$). All of the intervention contrasts were statistically significant but none of the control group contrasts were. The key contrasts, however, compare the improvements for the intervention group

against the improvements for the control groups by differencing them. Here are these results from the `MODEL CONSTRAINT` command syntax:

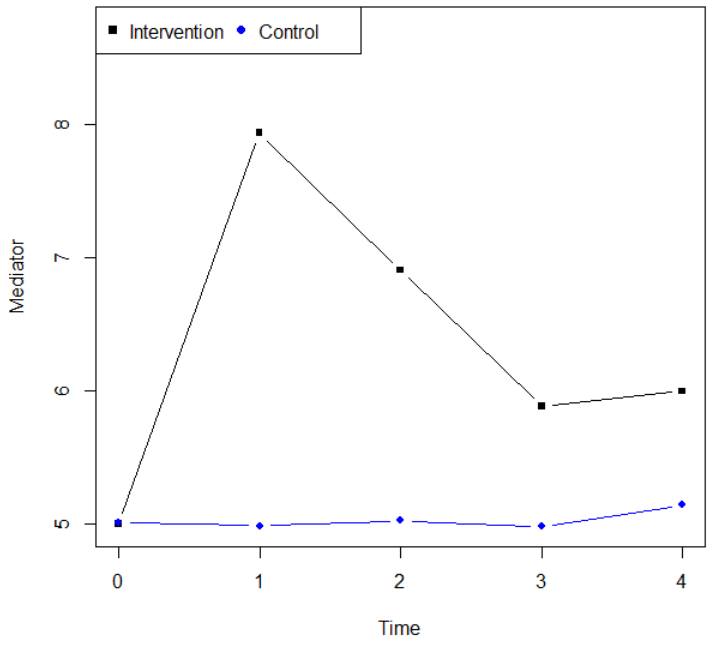
```
MODEL RESULTS

                                                    Two-Tailed
                  Estimate       S.E.  Est./S.E.    P-Value

New/Additional Parameters
    DIFFEY10         2.618      0.084     31.242      0.000
    DIFFEY20         1.849      0.105     17.585      0.000
    DIFFEY30         1.346      0.128     10.509      0.000
    DIFFEY40         1.473      0.145     10.139      0.000
```

The improvement from baseline to the immediate posttest for the intervention group was 2.58 ±0.12 but for the control group it was -0.04 ±0.12. The difference between these improvement rates was 2.62 ±0.17 (CR = 31.24, $p < 0.05$), an improvement of about 26 points on the original Y metric. The improvement from baseline to wave 4 for the intervention group was 1.52 ±0.20 but for the control group it was 0.05 ±0.22. The difference between these improvement rates was 1.47 ±0.29 (CR = 10.14, $p < 0.05$). Suppose that prior to the study, the project staff and research team decided than an improvement of 10 points or greater on the original 0 to 100 metric would be meaningful, which translates into a difference of 1.0 on the transformed metric. The 95% confidence interval for the wave 4 group difference contrasts was 1.18 to 1.76. Because the lower limit for this interval exceeds the meaningfulness standard, I declare this effect meaningful. This was true for all of the group difference contrasts.

In sum, the intervention had a meaningful effect at the immediate posttest. There was some decay in these effects over time but the intervention effect remained meaningful at each of the follow-ups.

Effect of the Treatment on the Mediator

To test the effects of the intervention on the mediator, I follow the same structure as that for the total effect on Y. Here is the plot for the model-implied mediator means at each of the time periods (see Lines 88-89 and 94-95 of the `MODEL CONSTRAINT` command):



The general patterning of the means is similar to that of the outcome. Here is the Mplus output that documents the improvement in the motivation to lead a healthy lifestyle at each time point relative to the baseline for each group:

```
MODEL RESULTS

                                                    Two-Tailed
                  Estimate       S.E.  Est./S.E.    P-Value

New/Additional Parameters

    EM10_I            2.936      0.056     52.585      0.000
    EM20_I            1.905      0.069     27.618      0.000
    EM30_I            0.886      0.083     10.629      0.000
    EM40_I            0.996      0.095     10.512      0.000

    EM10_C           -0.023      0.056     -0.417      0.677
    EM20_C            0.014      0.073      0.185      0.853
    EM30_C           -0.026      0.089     -0.289      0.773
    EM40_C            0.132      0.094      1.410      0.159
```

The significance tests for each row evaluate if the degree of improvement from baseline to the designated posttest/follow-up is different from zero. For example, the improvement in motivation from baseline to wave 4 for the intervention group was 1.00 ±0.19 which was statistically significantly different from zero (CR = 10.51, $p < 0.05$). All

of the intervention contrasts were statistically significant but none of the control group contrasts were. Here is the Mplus output that compares the improvement in motivation for the intervention group against the improvements for the control groups by differencing them:

```
MODEL RESULTS
                                                    Two-Tailed
                  Estimate       S.E.  Est./S.E.    P-Value

New/Additional Parameters

    DIFFEM10         2.959      0.079     37.538      0.000
    DIFFEM20         1.891      0.101     18.792      0.000
    DIFFEM30         0.911      0.122      7.481      0.000
    DIFFEM40         0.864      0.133      6.475      0.000
```

The improvement in motivation from baseline to the immediate posttest for the intervention group was 2.94 ±0.11 but for the control group it was -0.02 ±0.11. The difference between these improvement rates was 2.96 ±0.16 (CR = 37.54, $p < 0.05$), an improvement of about 30 points on the original M metric. The improvement in motivation from baseline to wave 4 for the intervention group was 1.00 ±0.19 but for the control group it was 0.13 ±0.19. The difference between these improvement rates was 0.86 ±0.26 (CR = 6.48, $p < 0.05$). Suppose that prior to the study, the project staff and research team decided an improvement of 10 points or greater on the original 0 to 100 metric of M would be meaningful. This translates into a difference of 1.0 on the transformed metric. The 95% confidence interval for the wave 4 group difference contrast was 0.60 to 1.12. Because this interval overlaps the meaningfulness standard, I cannot confidently conclude the effect is meaningful when taking sampling error into account. Given that the wave 4 effect was statistically significant ($p< 0.05$), I can confidently conclude the effect is non-zero. However, I cannot confidently conclude it is meaningful. This dynamic also was the case for the wave 3 data, but the results for wave 1 and wave2 were both statistically significant and meaningful.

In sum, the intervention had a meaningful effect on motivation at the immediate posttest. There was some decay in these effects over time but the intervention effect remained meaningful at wave 2. Perhaps a short intervention booster session 6 months after the intervention would be helpful.

Effect of the Mediator on the Outcome

For the multi-group model, the analysis estimates the effect of changes in the mediator on changes in the outcome separately for the intervention and control groups. In addition,

the coefficients are estimated separately at each wave. Here are the relevant coefficients:

```
                                                    Two-Tailed
                   Estimate       S.E.  Est./S.E.    P-Value

Group CONTROL

 DY1      ON
    DM1                0.331      0.044      7.498      0.000

 DY2      ON
    DM2                0.331      0.047      7.069      0.000

 DY3      ON
    DM3                0.375      0.043      8.764      0.000

 DY4      ON
    DM4                0.350      0.044      7.982      0.000


Group INTERVENTION

 DY1      ON
    DM1                0.302      0.045      6.756      0.000

 DY2      ON
    DM2                0.358      0.040      8.837      0.000

 DY3      ON
    DM3                0.342      0.041      8.336      0.000

 DY4      ON
    DM4                0.291      0.042      6.965      0.000
```

All of the coefficients are statistically significant and in the range of 0.30, meaning that for every one unit that motivation improves, healthy behaviors improve by about 0.30 units. On their original metrics, for every 10 units that motivation improves, health behavior is predicted to improve by about 3 units.

In the `MODEL CONSTRAINT` command, I tested if the values of the coefficient at each wave were statistically significantly different from one another (see Lines 76 to 83). Here are the results for the intervention group (where the last three characters of `1V2` in the label `TI1V2` reflects the M→Y coefficient at wave 1 minus the M→Y coefficient at wave 2; `11V3` in the label `TI1V3` reflects the M→Y coefficient at wave 1 minus the M→Y coefficient at wave 3, and so on):

```
MODEL RESULTS

                                                    Two-Tailed
                  Estimate       S.E.  Est./S.E.    P-Value

New/Additional Parameters

    TI1V2            -0.056      0.060     -0.943      0.345
    TI1V3            -0.041      0.061     -0.664      0.506
    TI1V4             0.010      0.058      0.177      0.859
    TI2V3             0.016      0.056      0.280      0.779
    TI2V4             0.066      0.055      1.202      0.229
    TI3V4             0.051      0.058      0.879      0.380
```

None of the pairwise contrasts were statistically significant. This also was true for the control group:

```
                                                    Two-Tailed
                  Estimate       S.E.  Est./S.E.    P-Value

New/Additional Parameters

    TC1V2             0.000      0.064      0.007      0.995
    TC1V3            -0.044      0.062     -0.712      0.476
    TC1V4            -0.019      0.063     -0.303      0.762
    TC2V3            -0.045      0.062     -0.725      0.468
    TC2V4            -0.019      0.063     -0.309      0.757
    TC3V4             0.025      0.062      0.408      0.683
```

I also tested if the M→Y coefficient at a given wave was statistically significantly different in the intervention group than the corresponding M→Y coefficient in the control group (see Lines 73 to 75 in the `MODEL CONSTRAINT` commands). Here are the results, where `G_DIFF` in the label stands for "group difference" and the number at the end refers to the wave at which the difference was evaluated):

```
                                                    Two-Tailed
                  Estimate       S.E.  Est./S.E.    P-Value

New/Additional Parameters

        G_DIFF1          -0.030      0.063     -0.470      0.638
        G_DIFF2           0.027      0.062      0.439      0.661
        G_DIFF3          -0.033      0.059     -0.560      0.575
        G_DIFF4          -0.059      0.061     -0.969      0.333
```

None of these contrasts were statistically significant. Given the above and the obvious homogeneity of the coefficients, it would not be unreasonable to refit the model but to impose equality constraints on the coefficients across time and across groups. I accomplish this in the main syntax (at Lines 30 to 33 and Lines 39 to 42) by assigning a common label to each coefficient. In the `MODEL intervention` commands I replace Lines 30 to 33 with

```
dy1 ON dm1 (by);

dy2 ON dm2 (by);
dy3 ON dm3 (by);
dy4 ON dm4 (by);
```

and in the `MODEL intervention` commands I replace Lines 39 to 42 with

```
dy1 ON dm1 (by);
dy2 ON dm2 (by);
dy3 ON dm3 (by);
dy4 ON dm4 (by);
```

I removed the `MODEL CONSTRAINT` commands and re-ran the syntax and obtained a chi square statistic for the constrained model of 43.36 with 45 degrees of freedom. This statistic can be compared to the original chi square with the unconstrained coefficients of 40.11 with 38 degrees of freedom. The difference between these indices is trivial and statistically non-significant by means of a chi square difference test (see Chapter 8). The estimated value of the constrained coefficient was 0.335 ±0.032 (CR = 21.00, $p < 0.05$). It provides a single best estimate of the effect of changes in the mediator on changes in the outcome. Suppose prior to the study the program staff deemed that a coefficient of 0.25 or larger would be meaningful. The 95% confidence interval for the above coefficient was 0.30 to 0.37. Because the lower limit of this interval exceeds the meaningfulness standard, the M→Y link is deemed meaningful in magnitude. Had the coefficients been non-trivially heterogenous, I would need to respect that heterogeneity and delve into the heterogeneous differences in more depth.

Omnibus Mediation Analysis

As noted throughout this chapter, omnibus mediation tests are of lower priority for me. I instead prefer to focus on the individual links in mediational chains, evaluating their effects sizes. If I want an overall omnibus test of statistical significance, I rely on the joint significance test, which in the current case affirms mediation because both the T→M link and the M→Y link were statistically significant ($p < 0.05$). The product coefficient

method is complicated by the fact that the T→M link is across-person in nature but the M→Y link is within-person in nature.

Concluding Comments on RET for Latent Change Score Example

In sum, the LCSM model suggest that the intervention had a meaningful overall effect on healthy behaviors at each time point and that it had a nonzero and meaningful effect on the motivation to lead a healthy lifestyle at the immediate posttest and first 3 month follow-up. The significant second and last 3 month follow-up, the effect remained statistically significant, but I could not confidently conclude that it had a meaningful effect. Finally, the mediator had a statistically significant and meaningful effect on the outcome for both the intervention and control groups and at each wave of the post-intervention data.

Although I did not report them above, here are the squared multiple correlations for the intervention and control groups for the phantom Y latent variables:

```
R-SQUARE

Group CONTROL

     Latent                                           Two-Tailed
    Variable        Estimate       S.E.  Est./S.E.    P-Value

    DY1                0.098      0.025      3.946      0.000
    DY2                0.095      0.025      3.748      0.000
    DY3                0.116      0.025      4.567      0.000
    DY4                0.101      0.025      4.105      0.000

Group INTERVENTION

     Latent                                           Two-Tailed
    Variable        Estimate       S.E.  Est./S.E.    P-Value

    DY1                0.079      0.023      3.498      0.000
    DY2                0.121      0.025      4.829      0.000
    DY3                0.104      0.024      4.364      0.000
    DY4                0.074      0.020      3.636      0.000
```

Overall, the improvement in the latent phantom mediator accounted for about 10% of the variation in the latent phantom outcome. This is non-trivial but it is clear from the 90% or so unexplained variances that there are other variables operating to impact health behavior improvement that the program designers should attend to in the future. Additional perspectives on this matter are provided by analyzing the direct effect of the

treatment on the phantom latent improvements independent of the phantom latent mediator, a topic I consider below.

In addition to the above analyses, I would repeat them using bootstrapping for sensitivity purposes and pursue standard outlier analyses and measurement error sensitivity analyses, per my standard practice. I also can explore supplemental indices of effect size, such as the probability of exceptions to the rule, described in Chapter 10.

Additional Issues for Latent Change Score Models

Estimating Direct Effects of the Treatment on the Outcome in Multi-group Modeling

In multi-group models where the treatment condition defines the grouping variable, the direct effect of the treatment condition on the outcome independent of the mediators is not obvious on Mplus output. I discuss how to isolate such effects in depth in Chapter XX. For the present example, consider the case where the latent change variable `dy1` is predicted from the latent change variable `dm1`. Here are the equations observed in each group for the constrained solution I pursued earlier:

`dy1_i` = 1.59 + 0.34 `dm1_i`

`dy1_c` = -0.03 + 0.34 `dm1_c`

The `i` signifies the intervention group and the `c` signifies the control group. It turns out that the direct effect of the treatment condition on `dy1` is captured by the group difference in the intercepts, namely (1.59) – (-0.03) = 1.62. It is the estimated effect of the treatment condition on `dy1` when `dm1` is held constant. I found that it was statistically significant ($p < 0.05$) when I added the contrast to the `MODEL CONSTRAINT` commands. In the current LCSM example, there are four equations for each group, one at each post-intervention wave, `dy1 ON dm1`, `dy2 ON dm2`, `dy3 ON dm3`, and `dy1 ON dm1`. This means there are four distinct intercept comparisons and, by implication, four direct effects of the treatment condition on the outcome. LCSMs are built to capture evolving processes over time by examining change on a granular basis from one time period to the next. Intercepts often vary across time in LCSMs to capture non-linear average trajectories of change over time. Allowing intercepts to vary freely across time allows the analyst to see *when* intervention and control group change trajectories diverge after controlling for M. This is why there are four distinct direct effects, not one.

Here is the direct effect result for `dy4` using the constrained solution:

`dy4_i` = 0.04 + 0.34 `dm4_i`

`dy4_c` = -0.10 + 0.34 `dm4_c`

such that (0.04) – (-0.10) = 0.14, a result that differs from that for `dy1`.

Measurement Error and Multiple Indicator Models

The current example used single indicators of Y and M. It is fairly straightforward to extend the model to analyze multiple indicators at each time point to better account for measurement error (Estrada et al., 2019; McArdle & Hamagami, 2004). A LCSM with multiple indicators patterned after Figure 16.28 is shown in Figure 16.29. Again, to reduce clutter in the diagram I omit disturbances and correlations that would be included in the Mplus program but I show the measurement errors. Each latent Y has three indicators. Only the Y variables have multiple indicators but it is straightforward to extend the model to the case where the M also have multiple indicators.

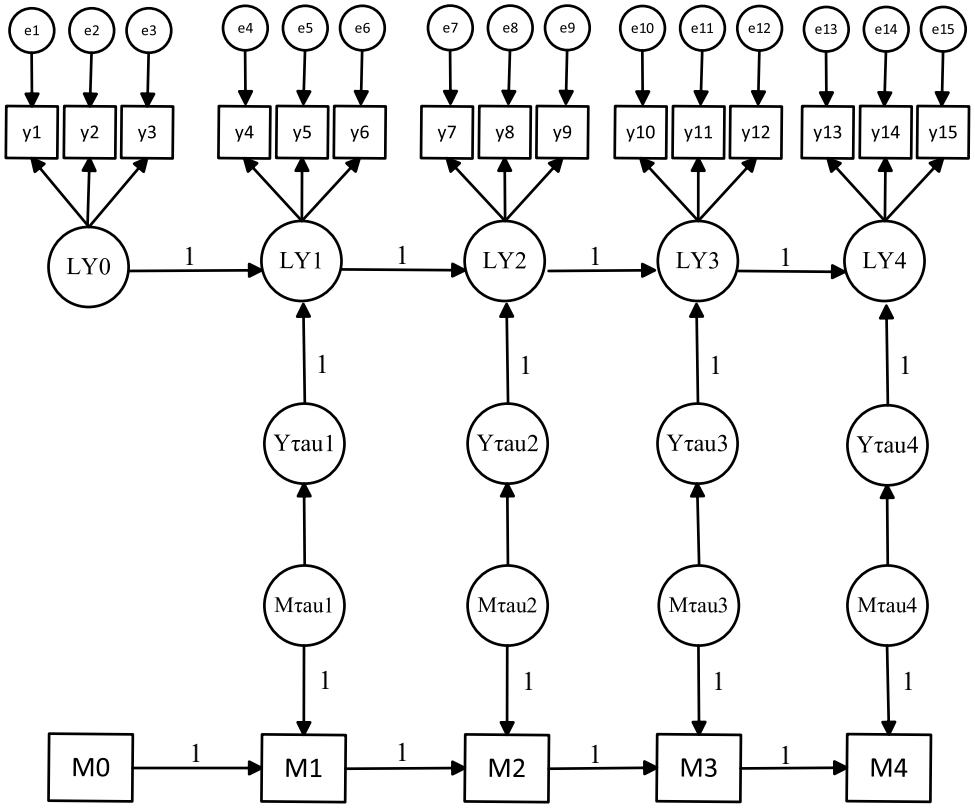
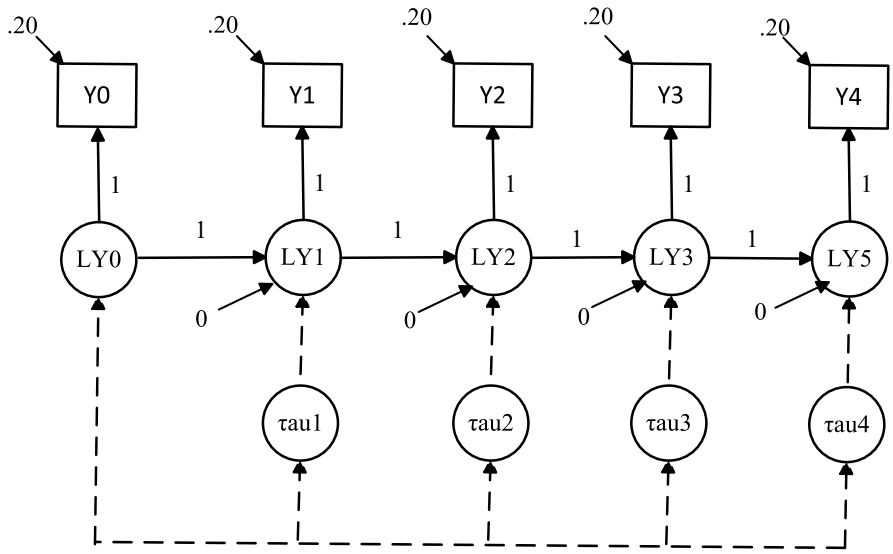


FIGURE 16.29 LCSM model with multiple indicators

In multiple indicator LCSMs, you select one of the observed variables to be the marker indicator for a given LY and then fix its measurement intercept to zero to pass its mean to the latent Y. You estimate the measurement intercepts for the two other indicators. You fix the intercepts of the Yamagami latent Lys to zero and their disturbance variances to be zero. Alternatively, with four or more waves, you can estimate them if you place an equality constraint on them. For additional details on such models, see Newsom (2023). With multiple indicator LCSMs, you should address issues of measurement invariance across groups and times (see the document on *testing measurement non-invariance* on the *Resources* tab of my webpage in Chapter 3). It also is possible to *adjust single indicator models for measurement error* using strategies discussed in the document on the *Resources* tab of my webpage for Chapter 3. See also the strategies discussed by Geiser (2021). Using the concepts in my web document and, to keep things simple, if I programmed the model in Figure 16.27, it would appear as follows:



Each of the observed variables is converted to a latent variable with a single indicator with a loading fixed to 1.0 and error variance fixed to my best guess of its unreliability variance, in this case 0.20. I also fix the measurement intercepts of the observed Y to 0. For the latent Ys, I fix their disturbance variances to zero as well as their measurement invariances. All of this is explained in the aforementioned document for Chapter 3. The strategy takes into account and adjusts for measurement error in the change scores.[16]

[16] As before, you can estimate the unreliabilities if you have more than four waves with equality constraint on them.

Concluding Comments on Latent Change Score Models for RETs

Latent change score models for the analysis of RETs have several advantages the most noteworthy of which is the granular analysis of change dynamics over time. Allison (1990) has described scenarios where a focus on change scores is preferable to regression modeling that covaries the baseline measure. LCSMs fit this scenario well. Having said that, LCSMs have multiple drawbacks. First, when working with more than one mediator the models can become quite complex and convergence within SEM software can be challenging. Model complexity and the number of parameters to be estimated becomes that much more challenging when you also work with multiple indicator models. LCSMs can produce bias in parameter estimates if you seek to control the baseline measure of the composite change scores (see Sorjonen et al., 2023). As a result, one often has less statistical power when using LCSMs than the other approaches I have discussed in this chapter. LCSMs also can be less efficient for dealing with regression to the mean phenomena and handling cross-lagged effects, autoregressive paths, and concurrent and lagged mediation across the multiple latent factors is a nontrivial exercise.

As you explore the method in the literature, you will find programming strategies for LCSMs can be vastly different, in part, because there are different traditions surrounding their use. Covariate control is relatively straightforward and follows the same principles as those discussed in prior sections of this chapter. For a discussion of applying the method to binary and ordinal outcomes, see Newsome (2023).

MIXED MODELS AND RETs

A popular strategy in the social and health sciences for RCT analysis with longitudinal data is commonly referred to as **mixed modeling** or **linear mixed modeling**. The strategy is called "mixed" because modeling includes both fixed and random effects/coefficients, although the effects or coefficients declared as random or fixed vary from one study to another. The term *mixed model* is an umbrella term that is ambiguous because it is instantiated in so many different forms. Sometimes the technique is used to refer to traditional repeated measures analysis of variance described in classic experimental design books (Maxwell, Delaney & Kelley, 2017; Kirk 2012). Other times it refers to latent curve modeling and other times it refers to specialized forms of hierarchical linear modeling (HLM). The classic text by Singer and Willitt (2003) often is cited as a source for exposition of mixed models for longitudinal data. It is an excellent resource but its fundamental focus is on multilevel modeling using long format data. Their methods have been improved upon by more recent SEM modeling strategies described in the current chapter and in Chapter 25.

A classic RCT example that applies multilevel mixed models to a randomized trial is by Brem, Shorey, Ramsey and Stuart (2023). They studied women exposed to interpersonal violence (IPV) and whether adding an alcohol intervention to an intervention for battered women decreases women's alcohol use as well as IPV outcomes. Women were randomized to the standard intervention condition versus an augmented intervention that included an alcohol target module. The primary outcome variables were the percentage of days abstinent from alcohol (PDAA) and the percentage of heavy drinking days (PHDD) over the past 6 months when measured at baseline and since the time of the prior interview at post-intervention. Data were collected at baseline and 3-, 6-, and 12-month follow-ups. Analyses took the form of a multilevel model with time nested within individuals. In long format, the data were structured as follows (shown for the first 3 individuals (where standard intervention = 0, augmented intervention= 1:[17]

ID	TIME	PDAA	PHDD	GROUP
1	0	58.3	25.9	0
1	3	70.2	18.3	0
1	6	72.7	15.4	0
1	12	77.5	13.2	0
2	0	60.0	8.5	1
2	3	77.6	10.6	1
2	4	81.1	10.7	1
2	12	81.8	10.8	1
3	0	59.9	24.9	0
3	3	71.2	19.2	0
3	6	71.7	14.4	0
3	12	71.5	14.0	0

The researchers used maximum likelihood multilevel modeling to analyze the data with individuals defined as the between-person source (called *Level 2*) and time as the within-person source (called *Level 1*). They created a product term between time and group (GrpXTime) and then estimated a model of the form $Y = a + b_1$ Time $+ b_2$ Group $+ b_3$ GrpXTime), taking into account the dependency structures for the Level 1 data and treating the Level 1 intercepts as random. The Y in one analysis was PDAA and in another analysis it was PHDD. The researchers predicted a statistically significant coefficient for the interaction term because the effect of the intervention (Group) was thought to vary depending on the time at which the outcome variable was measured (baseline versus 3 months versus 6 months versus 12 months). The data were analyzed using the software package *HLM* (Raudenbush et al., 2011). The researchers decomposed the interaction effect to provide insights into intervention effects at each time point.

[17] The values I provide are hypothetical but representative of the study results.

This mixed model approach for RCTs is not unreasonable for outcome-only studies. However, such models are not amenable to RETs where mediators are part and parcel to the analysis. The traditional mixed model approach dos not have the flexibility of the SEM models described in this chapter nor can they easily deal with latent variables, measurement error, and complex relationships among mediators and disturbances. The version I illustrate in this chapter uses Bayes estimation which is capable of addressing such matters while avoiding numerical integration. It also addresses Nichol and Ludtke bias and handles missing data better than maximum likelihood and multiple imputation methods. Imputation methods can break the nested data structure producing biased standard errors. The *MICE* package in R offers a specialized multi-level chained equation approach that uses a multilevel model as the imputation engine for each missing variable. However, it lacks the elegance of Bayesian-based multi-level modeling as implemented in Mplus. Mplus Bayes fully integrates missing data into the model analysis whereas MICE instead relies on a separate sequential imputation phase.

The hypothetical data example I use maps onto the structure of Brem et al. (2023) with measures of a mediator M and an outcome Y obtained at baseline and at 3, 6, and 12 month follow-ups. M and Y represent total scores on multi-item scales that are the average of item-level responses, hence they are treated as continuous. The standard deviations for M and Y are near 1.0, which gives you a sense of their metrics (i.e., they are much like standard score metrics). Y is an index of healthy exercising and M is an in index of the general motivation to engage in healthy exercise. The intervention sought to increase M on the assumption it would translate into increases in Y. The data were modeled so as to include a direct effect from the intervention to Y independent of M. The data need to be in long format for the Bayes method I illustrate.

The interaction term in the Brem et al. (2023) study presumes an interaction of a specific form, namely a bi-linear interaction (see Chapter 18). This often is unrealistic in RETs where there often is a spike in change upward from baseline to the first post-intervention assessment followed by intervention decay over time after that, decay that can be uneven. I therefore represented time with dummy variables that I calculated from the time variable in the data set. The dummy variables are called t1, t2 and t3 with the baseline assessment serving as the reference period. This allows me to accommodate a wide range of possible decay functions. Table 16.21 presents the Mplus syntax.

Table 16.21. Mplus Syntax for Mixed Model Example

```
1.  TITLE: Mixed model analysis with mediation ;
2.  DATA: FILE = mixedM.dat ;
3.  VARIABLE:
4.    NAMES = id time m y treat t1 t2 t3;
```

```
5.    USEVARIABLES = m y treat t1 t2 t3;
6.    CLUSTER = id;
7.    BETWEEN = treat;
8.    WITHIN = t1 t2 t3;
9.  ANALYSIS:
10.   TYPE = TWOLEVEL RANDOM;
11.   ESTIMATOR = BAYES; FBITERATIONS = 26700 ;
12. MODEL:
13.  %WITHIN%
14. !Latent random slopes/coefficients tracking mediator over time
15.   s_m1 | m ON t1;
16.   s_m2 | m ON t2;
17.   s_m3 | m ON t3;
18. !Within-person structural mediation path with label b
19.   y ON m (b);
20. !Latent random slopes/coefficients tracking outcome paths
21.   s_y1 | y ON t1;
22.   s_y2 | y ON t2;
23.   s_y3 | y ON t3;
24. !Estimate variances
25.   m; y;
26.  %BETWEEN%
27. !Regress the random coefficient variables on treatment status
28.   s_m1 ON treat (a1);
29.   s_m2 ON treat (a2);
30.   s_m3 ON treat (a3);
31.   s_y1 ON treat (dir_t1);
32.   s_y2 ON treat (dir_t2);
33.   s_y3 ON treat (dir_t3);
34. !Below are treatment effects at baseline (Time 0); imbalance checks
35.   m ON treat (bg_m);
36.   y ON treat (bg_y);
37. !Baseline intercepts for the control group
38.   [m] (int_m);
39.   [y] (int_y);
40. !Time trajectories for the control group (natural drift)
41.   [s_m1] (c_m1);  [s_m2] (c_m2); [s_m3] (c_m3);
42.   [s_y1] (c_y1);  [s_y2] (c_y2); [s_y3] (c_y3);
43. !Freely estimate all variances and covariances
44.   m y s_m1 s_m2 s_m3 s_y1 s_y2 s_y3;
45.   m y s_m1 s_m2 s_m3 s_y1 s_y2 s_y3 WITH m y s_m1 s_m2 s_m3 s_y1 s_y2 s_y3;
46. !Allow baseline covariance for M and Y
47.   m WITH y;
48. MODEL CONSTRAINT:
49. !Estimated cell means for mediator
50.   NEW(m_ctrl0 m_ctrl1 m_ctrl2 m_ctrl3 m_trtx0 m_trtx1 m_trtx2 m_trtx3);
51. !Control group M means/medians
52.   m_ctrl0 = int_m;
53.   m_ctrl1 = int_m + c_m1;
54.   m_ctrl2 = int_m + c_m2;
55.   m_ctrl3 = int_m + c_m3;
56. !Intervention group M means
```

```
57.   m_trtx0 = int_m + bg_m;
58.   m_trtx1 = int_m + bg_m + c_m1 + a1;
59.   m_trtx2 = int_m + bg_m + c_m2 + a2;
60.   m_trtx3 = int_m + bg_m + c_m3 + a3;
61. !Estimated cell means for outcome
62.   NEW(y_ctrl0 y_ctrl1 y_ctrl2 y_ctrl3 y_trtx0 y_trtx1 y_trtx2 y_trtx3);
63. !Control group Y means
64.   y_ctrl0 = int_y + b*m_ctrl0;
65.   y_ctrl1 = int_y + c_y1 + b*m_ctrl1;
66.   y_ctrl2 = int_y + c_y2 + b*m_ctrl2;
67.   y_ctrl3 = int_y + c_y3 + b*m_ctrl3;
68. !Intervention group Y means
69.   y_trtx0 = int_y + bg_y + b*m_trtx0;
70.   y_trtx1 = int_y + bg_y + c_y1 + dir_t1 + b*m_trtx1;
71.   y_trtx2 = int_y + bg_y + c_y2 + dir_t2 + b*m_trtx2;
72.   y_trtx3 = int_y + bg_y + c_y3 + dir_t3 + b*m_trtx3;
73. !Total effects of treatment on Y
74.   NEW(tot_eff1 tot_eff2 tot_eff3);
75.   tot_eff1 = (y_trtx1 - y_trtx0) - (y_ctrl1 -  y_ctrl0 )   ;
76.   tot_eff2 = (y_trtx2 - y_trtx0) - (y_ctrl2 -  y_ctrl0 )   ;
77.   tot_eff3 = (y_trtx3 - y_trtx0) - (y_ctrl3 -  y_ctrl0 )   ;
78. !Indirect Effects
79.   NEW(ind_t1 ind_t2 ind_t3);
80.   ind_t1 = a1 * b;
81.   ind_t2 = a2 * b;
82.   ind_t3 = a3 * b;
83. !Decay contrasts
84.   NEW(ydecay12 ydecay23 mdecay12 mdecay23);
85. !Tests if the total effect drops from Time 1 to Time 2
86.   ydecay12 =  tot_eff2 - tot_eff1 ;
87. !Tests if the total effect drops from Time 2 to Time 3
88.   ydecay23 =  tot_eff3 - tot_eff2 ;
89. !Tests if the effect on mediator drops from Time 1 to Time 2
90.   mdecay12 =  a2 - a1 ;
91. !Tests if Effects effect on mediator drops from Time 2 to Time 3
92.   mdecay23 =  a3 - a2 ;
93. OUTPUT: STDYX CINTERVAL(HPD) RESIDUAL TECH4 TECH8;
```

Line 6 uses the respondent ID to define each individual as a separate cluster. Line 7 identifies for Mplus the variables that are to be exclusively between-person in character (and not within-person); Line 8 identifies for Mplus the variables that are exclusively within-person in character. I list the three time dummy variables on this line to indicate they will only be used in a within-person capacity. Line 10 indicates the analysis will be multilevel in character (time nested within individuals) and that there will be at least one random coefficient in the model. Line 11 tells Mplus to use Bayes estimation with a fixed number of iterations set to 26,700. The `FBITERATIONS` subcommand causes Mplus to disable its internal Gelman-Rubin convergence check based on PSRs and forces every chain to execute for exactly 26,700 iterations. I chose this value for the current analysis to

facilitate efficient convergence should you choose to run the program on your computer. Sometimes convergence is slow and can take hours with complex multilevel SEM. I spare you that misery by informing you that if you use 26,700 iterations, convergence will be achieved and you do not need to let Mplus iterate beyond that. Note that in practice if you use `FBITERATIONS` you must verify satisfactory convergence by checking diagnostics generated by `TECH8` on the output and other diagnostics.

Lines 13 to 25 specify the within-person model to estimate. It is here where I define the coefficients I want to treat as random. Lines 15 -17 allow the T→M effect to vary across individuals because the coefficients for it are treated as random. I now explain the syntax format of these lines. The entry to the left of the pipe | is a label I assign to the coefficient specified in the regression expressed to the right of the pipe. For Line 15, I regress the mediator onto the treatment dummy variable `t1` and give the resulting coefficient(s) the label `s_m1`; for Line 16, I regress the mediator onto the treatment dummy variable `t2` and give the coefficient(s) the label `s_m2`; for Line 17, I regress the mediator onto the treatment dummy variable `t3` and give the coefficient(s) the label `s_m3`. Although the three dummy variable regressions are listed on separate lines, Mplus treats them multivariately and regresses the mediator onto all of the predictors simultaneously. Given this, the reference group for each coefficient is the mediator measured at baseline. The coefficient `s_m1` thus represents the difference between the mediator at time 1 minus the mediator at time 0 (the baseline); the coefficient `s_m2` represents the difference between the mediator at time 2 minus the baseline mediator; the coefficient `s_m3` represents the difference between the mediator at time 3 minus the baseline mediator.

Line 19 tells Mplus to estimate the effect of M on Y on a within person basis. This coefficient is treated as fixed, that is, it is assumed to take on the same value for each individual. Some researchers might decide to treat being random coefficient in character, but in this case, the research team felt the assumption of a fixed effect is reasonable based on substantive considerations. One needs to be careful about willy-nilly making coefficients random because if you have many such effects it can lead to convergence difficulties and oppressive computational demands. Also, the fixed coefficient parameterization represents a more parsimonious model.

Lines 20-21 add to the prediction of Y at the within-person level the direct effects of the treatment on Y independent of the mediator. These direct effects are treated as having random coefficients. Finally, Line 15 estimates the disturbance variances for the two endogenous variables in the within-person model, M and Y.

Lines 26 to 47 reflect the between-person model. Lines 28 to 33 regress the random coefficients defined in the within-person models onto the two group intervention versus

control dummy variable called `treat`. Lines 35 and 36 check for baseline imbalance on the mediator and the outcome. The underlying logic for why this is the case goes something like this: When I regress the mediator M and outcome Y onto the time dummy variables t1, t2, t3 in the `%WITHIN%` section, Mplus automatically moves the random intercepts for these regressions to the `%BETWEEN%` level and represents them simply as M and Y at this level. These intercepts capture individual variations in the reference category at the within-person level which turns out to be in the current case the baseline M and Y. As such, specifying `m ON treat` and `y ON treat` at the between-person level directly tests for treatment group differences at time 0 and provides perspectives on the success of the randomization process.

Lines 37 to 42 estimate means/intercepts for key variables in the analysis and Lines 43 to 47 estimate key variances and covariances, including the covariances between the random coefficients. The intercepts for the random slopes/coefficients (bracketed as [s_m1] through [s_y3]) on Lines 41 and 42 represent shifts in control group means over time, which is sometimes referred to as **natural longitudinal drift**. It reflects changes in means across time that occur naturally absent the intervention. We ultimately want to compare such shifts to those in the intervention group.

Note that throughout the syntax I assign labels to many of the parameters because I Effects make use of those parameters in the `MODEL CONSTRAINTS` command.

In the `MODEL CONSTRAINTS` command, I specify different contrasts and parameters that I want Mplus to estimate that are of substantive interest. In general, these contrasts and parameters follow the underlying logic of Appendix A, although Appendix A focuses on latent growth curve modeling. However, the basic principles being used apply here. Lines 52 to 72 estimate the model-implied mediator and outcome means for the two treatment groups at each of the study assessment periods. Lines 73 to 77 estimate the total effect of the treatment condition on the outcome at each post-intervention assessment. Lines 78 to 82 estimate the indirect effect of the treatment condition on the outcome at each assessment point. Lines 83 to 92 estimate decay effects that I elaborate later.

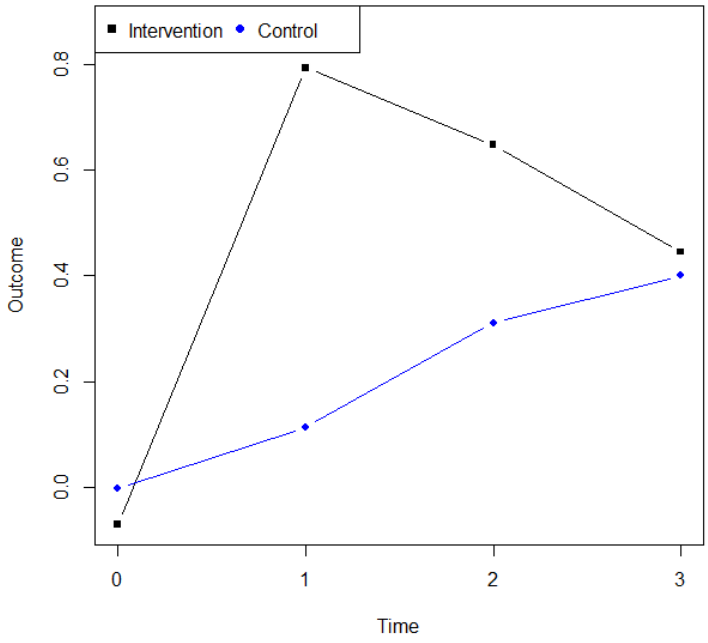
Like all SEM models, the Bayesian estimates of group central tendencies and path coefficients are constrained by the posited SEM model and the underlying Bayesian algorithms. Statistically, for point estimates of means, Bayes uses the median of the estimated posterior distributions for means and this also is true of path coefficients.

Upon execution of the syntax, the initial results suggested reasonable model fit. In the interest of space I do not review the fit indices here. The analytics follow closely those described for evaluating Bayesian model fit for the section on dynamic structural equation modeling. I now turn to the three core questions for an RET focused on mediation: (1) is there an overall meaningful effect of the intervention on the outcome,

(2) is there a meaningful effect of the intervention on the presumed mediator(s), and (3) does the mediator have a meaningful effect on the outcome.

Effect of the Intervention on the Outcome

To provide context, it is helpful to plot the outcome estimated means (from the `MODEL CONSTRAINTS` command) for the intervention and control groups at each of the time points. I used the Temporal line plot program on my website to create the plot:



There are several noteworthy trends. At baseline (Time 0), the intervention and control groups have roughly equal Y means, which is consistent with randomization. There is a clear-cut intervention effect relative to the control group at Time 1. The change from baseline is much greater in the intervention condition than the control condition. The control group exhibits natural longitudinal drift in that the estimated mean Y shows a steady increase over time. This underscores the importance of control group inclusion in randomized trials. Without the control group, what was virtually complete intervention decay by Time 3 would mistakenly appear to be only partial decay.

In the current study, there is not a single "total effect" but instead three "total effects," one at each of the post-intervention assessments. The treatment and control differences are from the `MODEL CONSTRAINT` section and appear as follows (edited):

	Estimate	Posterior S.D.	95% C.I. Lower 2.5%	Upper 2.5%	Sig
New/Additional Parameters					
TOT_EFF1	0.751	0.118	0.520	0.982	*
TOT_EFF2	0.403	0.115	0.177	0.628	*
TOT_EFF3	0.114	0.113	-0.109	0.335	

For the 3 month "immediate" posttest (Time 1), the estimated mean difference in Y for the intervention condition mean at the immediate posttest minus the intervention baseline mean was (0.792)-(-0.071) = 0.863. The corresponding value for the control group was (0.113)-(-0.002) = 0.115. The difference between these two differences, 0.863-0.115, is 0.75, which is the total effect at month 3. In this formulation, the total effect takes on a difference-in-difference (DiD) parameterization that is typically used in economics (e.g., Angrist & Pischke, 2009).[18] The 95% credible interval for the total effect was 0.52 to 0.98. Because the credible interval does not contain the value 0, the total effect is declared to be statistically significant ($p < 0.05$).[19] Suppose the program staff and research team decided *a priori* that a meaningful effect was a difference greater than 0.33. Because the lower limit of the credible interval was greater than 0.33, the total effect at the first follow-up is declared to be meaningful.

For the total effect at 6 months (Time 2), the estimated mean difference in Y for the intervention condition mean minus the intervention baseline mean was (0.648)-(-0.071) = 0.719. The corresponding value for the control group was (0.311)-(-0.002) = 0.313. The difference between these two differences is 0.41, which is the total effect at month 6. The 95% credible interval for the total effect was 0.18 to 0.63. Because the credible interval does not contain the value 0, the total effect is declared to be statistically significant ($p < 0.05$), i.e., it can confidently be said to be non-zero in the population. However, using the same meaningfulness standard, the credible interval overlaps the standard so we cannot confidently conclude the effect is meaningful. To be sure, the point estimate is suggestive of a meaningful effect but when we take into account the broader posterior distribution of the effect, we can't confidently characterize the effect as meaningful.

For the total effect at 12 months (Time 3), the estimated mean difference in Y for the intervention condition mean minus the intervention baseline mean was (0.445)-(-0.071) = 0.516. The corresponding value for the control group was (0.400)-(-0.002) =

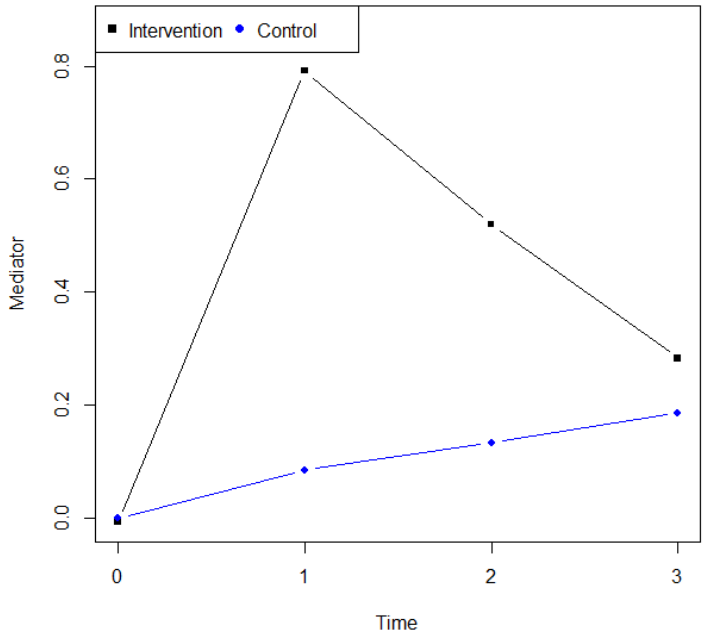
[18] DiD approaches are typically used in quasi-experiments but they can be used in randomized trials. I used this particular parameterization because it is common in traditional mixed modeling of RCTs. I could instead parameterize the model using the baseline Y as a covariate but doing so is beyond the scope of this chapter.

[19] The estimated means I used to calculate the respective differences in the calculations derive from Lines 62 to 72 of the MODEL CONSTRAINT command. The total effects are specified in Lines 75-77.

0.402. The difference between these two differences is 0.11, which is the total effect at month 12. The 95% credible interval for the total effect was -0.109 to 0.335. Because the credible interval contains the value 0, the total effect cannot be declared as "statistically significant ($p < 0.05$)". We can't confidently say the effect is non-zero in the population.

Effect of the Intervention on the Mediator

To provide context, I plot the mediator estimated means (from the `MODEL CONSTRAINTS` command) for the intervention and control groups at each of the time points. I used the Temporal line plot program on my website to create the plot:



The trends for the outcome also manifest themselves for the mediator. As before, there is not a single effect of the treatment condition on the mediator but instead three such effects, one at each of the post-intervention assessments. The treatment and control differences are taken from the `MODEL CONSTRAINT` commands and appear as follows:

```
Between Level

                                Posterior          95% C.I.
                  Estimate       S.D.     Lower 2.5%  Upper 2.5%    Sig

S_M1 ON TREAT        0.714       0.090        0.539       0.891       *
S_M2 ON TREAT        0.392       0.090        0.221       0.574       *
S_M3 ON TREAT        0.103       0.091       -0.069       0.288
```

For the 3 month posttest (Time 1), the estimated mean difference in M for the intervention condition mean at the "immediate" posttest minus the intervention baseline mean was (0.792)-(-0.006) = 0.798. The corresponding value for the control group was (0.085)-(0.000) = 0.085. The difference between these two differences, 0.798-0.085, is 0.71, which is the effect of the treatment condition on the mediator at month 3. As with the outcome, this is parameterized as a difference-in-difference. The 95% credible interval for the effect was 0.54 to 0.69. Because the credible interval does not contain the value 0, the effect is declared to be statistically significant ($p < 0.05$). Suppose the program staff and research team decided *a priori* that a meaningful effect was a difference greater than 0.33. Because the lower limit of the credible interval is greater than 0.33, the effect of the treatment on the mediator at the first follow-up is declared to be meaningful.

For the 6 month posttest (Time 2), the estimated mean difference in M for the intervention condition mean at the "immediate" posttest minus the intervention baseline mean was (0.520)-(-0.006) = 0.526. The corresponding value for the control group was (0.134)-(0.000) = 0.134. The difference between these two differences, 0.526-0.134, is 0.39, which is a difference-in-difference for the mediator. The 95% credible interval for the effect was 0.221 to 0.574. Because the credible interval does not contain the value 0, the effect is declared to be statistically significant ($p < 0.05$). Using the same meaningfulness standard, the credible interval overlaps the standard so we cannot confidently conclude the effect is meaningful. To be sure, the point estimate is suggestive of a meaningful effect but when we take into account the broader posterior distribution of the effect, we can't confidently characterize the effect as meaningful.

For the 12 month posttest (Time 3), the estimated mean difference in M for the intervention condition mean at the "immediate" posttest minus the intervention baseline mean was (0.283)-(-0.006) = 0.289. The corresponding value for the control group was (0.186)-(0.000) = 0.186. The difference between these two differences, 0.289-0.186, is 0.10, which, again, is a difference-in-difference for the mediator. The 95% credible interval for the effect was -0.069 to 0.288. Because the credible interval contains the value 0, the effect cannot be declared as "statistically significant ($p < 0.05$)." We can't confidently say the effect is non-zero in the population.

Effect of the Mediator on the Outcome

The estimated effect of the mediator on the outcome was treated as a fixed coefficient that has the same value for all individuals. The path coefficient for it is as follows:

```
Within Level

                        Posterior          95% C.I.
             Estimate       S.D.     Lower 2.5%  Upper 2.5%    Sig

Y ON M          0.482       0.229       0.067       0.956       *
```

For every one unit that M increases, Y is predicted to increase by 0.48 units. The credible interval was 0.067 to 0.956. Because it does not contain the value of 0, the null hypothesis of a zero coefficient is rejected, $p < 0.05$. Suppose prior to the study the program staff and research team felt a meaningful coefficient would be one of 0.3 or greater. Because the credible interval overlaps the standard so we cannot confidently conclude the effect is meaningful. To be sure, the point estimate is suggestive of a meaningful effect of M on Y but when we take into account the broader posterior distribution, we can't confidently characterize the effect as meaningful.

The above coefficient represents what is known as a **detrended** coefficient. It is detrended because time was treated as a covariate by virtue of including the three time dummy variables in the equation predicting Y from M. These dummy variables control for the overall growth or decline trajectories produced by time-varying variables external to the study, such as changes in seasons, downturns in the economy and the like. Evidence that external variables were operative is suggested by the longitudinal drift observed in the control group across time. The concern when such drift occurs is that the apparent effect of M on Y is inflated by uncontrolled time varying influences on both M and Y that serve as confounders.

Some methodologists argue that control of unmeasured time-varying confounds via the use of demeaning is scientifically more rigorous than not demeaning the data. Other methodologists argue against demeaning. The underlying logic of critics is that if a naturally occurring, external force causes M to rise over time, and if M then genuinely impacts Y, then that natural rise in M should cause a downstream rise in Y and taken into account accordingly. By stripping away the shared time trend, the argument goes, researchers are throwing the baby out with the bathwater and thereby underestimating the strength of the M→Y link. If the causal dynamic is a fundamental, invariant law of nature, it should operate regardless of what triggers the change in M. In this sense, the time-trend variance *is* genuine mediation variance and should not be stripped away.

I think that both sets of arguments have merit. A compromise to the controversy is to calculate the M→Y coefficient with and without detrending to determine if one's conclusions change across the two forms of analysis. In this particular study, the observed longitudinal drift in both M and Y in the control group is disconcerting. The de-trended coefficient provides a strict, conservative test of the RET specific M→Y

dynamic. The non-detrended coefficient provides the real-world total system association over time, recognizing that it likely includes both the causal effect as well as shared ambient trends.

I conducted an additional Mplus analysis to isolate the non-trended coefficient using the syntax in Appendix C. Here is the result:

```
Within Level

                    Posterior          95% C.I.
            Estimate       S.D.     Lower 2.5%  Upper 2.5%    Sig

 Y ON M         0.847       0.016       0.817       0.879        *
```

The estimated effect is much stronger and exceeds the meaningfulness standard of the prior analysis. I also regressed Y onto M at a between person-level in the syntax to determine if variation in people's average Y across time are linked to their average M across time. Here is the result for the analysis:

```
Between Level

                    Posterior          95% C.I.
            Estimate       S.D.     Lower 2.5%  Upper 2.5%    Sig

 Y ON M         0.855       0.074       0.719       1.001        *
```

The analyses converge with the detrended tests as to the statistical significance of the effect of M on Y. However, in the detrended analysis the effect did not meet meaningfulness standards whereas in the non-detrended analysis it did. Faced with such a scenario, my approach is to be transparent about matters but to try to resolve the disparity on substantive grounds. In practice many researchers do not de-trend but they instead include time-varying covariates that they believe may bias the M→Y link to bring such forces under control.

Direct Effect of the Intervention on the Outcome

The estimated effect of the intervention on the outcome independent of the mediator also is of interest. Here is the relevant Mplus output:

```
Between Level

                                    Posterior          95% C.I.
                        Estimate       S.D.      Lower 2.5%  Upper 2.5%    Sig

S_Y1 ON TREAT              0.399      0.189        0.032       0.766       *
S_Y2 ON TREAT              0.209      0.129       -0.044       0.462
S_Y3 ON TREAT              0.063      0.098       -0.130       0.254
```

The direct effects are evaluated separately at each time point. At the 3 month follow-up, the coefficient of 0.399 for the effect of the treatment on Y independent of M was statistically significant ($p < 0.05$) because the value of 0 was not contained in its credible interval (95% CI = 0.032 to 0.766). However, such was not the case for the 6 month and 12 month follow-ups. The statistically reliable effect at the 3 month follow-up suggests the effect of the intervention on Y cannot be fully accounted for by the target mediator, namely the motivation to engage in healthy exercise. The sole purpose of the intervention was to increase motivation with the idea that would increase healthy exercise. Perhaps a more fine grained analysis of program content and activities would reveal aspects of the program that touch on more than motivation per se and that are responsible for the program effects independent of motivation. produced independent effects on Y. Or perhaps something as simple as client liking of the program interventionists and a desire to please them may have impacted their behavior independent of facets of motivation that program content per se addressed. The fact that the direct effects did not persist over time suggests that the impact of these ancillary factors, whatever they are, diminished over time.

Omnibus Mediation

As noted, omnibus mediation indices are of lesser importance to me for purposes of program evaluation, but some researchers may find them to be of interest. The mediation analyses can be effectively explored at each separate time point using the joint significance strategy discussed in Chapter 9. In the current case, the M→Y link was statistically significant at all time periods but the be T→M link was statistically significant at times 1 and 2 but not time 2. This implies mediation at the earlier time lags but not at the 12 month follow-up. For the product coefficient method, I calculated in the `MODEL CONSTRAINT` commands the indirect effects at each time period on Lines 80 to 82. Here are the results:

```
                    Posterior          95% C.I.
             Estimate       S.D.      Lower 2.5%  Upper 2.5%    Sig

New/Additional Parameters

    IND_T1       0.339      0.170       0.020       0.689         *
    IND_T2       0.182      0.102       0.003       0.394         *
    IND_T3       0.042      0.055      -0.042       0.172
```

These results parallel those of the joint significance test.

Concluding Comments on Mixed Models

Mixed models are a popular method for analyzing RCT data, particularly in public health and medicine. I discussed earlier reasons that SEM-based implementations of them are preferred. The example I presented was simplistic because it included only a single mediator and no covariates. However, it is fairly straightforward to extend the programming strategies to such cases based on my discussion of covariate inclusion and multiple mediator cases in this chapter, although the `MODEL CONSTRAINT` section can become rather elaborate.

In some respects, the mixed model strategy and my programing approach to it can be framed via a 2 X 3 between-within factorial design, as follows:

	Baseline	Time 1	Time2	Time3
Treated Group				
Control Group				

We seek to analyze the outcome means and the mediator means in the context of this design as well as the M→Y link. Traditional repeated measures ANOVA is restrictive in this sense and for that matter so are traditional mixed modeling methods. The SEM based approach, by contrast, brings considerable flexibility to the task. In the `MODEL CONSTRAINT` commands on Lines 49 to 72, I estimate the cell means for the above design. I then use the cell means to address a variety of substantively interesting contrasts, such as the difference-in-difference approach to evaluating the treatment impact on the outcome, namely the total effects in Lines 73 to 77. Mplus provides the programming flexibility to explore virtually any contrast in the above table. For example, on Lines 83 to 88, I added formal contrasts to determine if the decay in the total effects for the outcome are statistically significant and to yield credible intervals for the

contrasts. The contrasts for the cell means also can be done with latent variables and covariates, both time-invariant and time-varying as dictated by the nature of the contrast.

An advantage of using the Bayes approach is that it handles missing data elegantly and by specifying HPD credible intervals on the `OUTPUT` line, you use the Bayesian version of bootstrapping for robustness. As a final note, my treatment of time as a dummy variable in the RET example complicated the programming strategy but I wanted to give you that flexibility because I think it is more realistic in applied settings.

ANALYSIS OF MIDTREATMENT EFFECTS

Mid-treatment assessments in clinical trials are common. In many medical, biological and pharmacological oriented trials, the mid-treatment assessments are used as check-ins to monitor adverse events and ensure patient safety. It is good methodological practice to pre-specify when these assessments are to be made, as appropriate. Treatment dropouts resulting from safety considerations are a potential source of bias to randomization. I discuss strategies for dealing with such bias in Chapter 27.

Mid-treatment assessments also take the form of pre-planned statistical evaluations of accumulated data at key study milestones. These evaluations help identify if a trial is overwhelmingly successful or if it is futile and should be terminated to save resources. Sometimes mid-treatment assessments are made to tweak or alter trial parameters, such as adjusting sample sizes or modifying the target population. Such decisions must be scientifically justified where possible. There are multiple resources for planning and implementing mid-treatment assessments which I provide on the *Resources* tab of my web page under Chapter 16.

My focus in this section is primarily on statistical issues for the analysis of studies that use mid-treatment for substantive rather than methodological or safety reasons and seek to make statements about intervention evaluation that takes the midtreatment assessment into account.

Early Responders

One use of mid-treatment assessments is to identify early responders to treatment with the idea that treatment can then be streamlined for them to conserve time, energy, and resources for all concerned. The practice also is important for adaptive designs and precision medicine, but it also introduces statistical challenges. One issue is that of regression to the mean: Individuals who show a strong early response often naturally drift back toward the population average over time or can do so because of the dynamics of

measure fallibility (see Chapter 4). Failing to account for this phenomenon can falsely identify early responders.

Another challenge occurs for complex multi-component interventions. Suppose a treatment for cognitive therapy emphasizes cognitive restructuring during the first few sessions, after which emotion regulation is addressed followed by behavioral activation and then exposure, i.e., facing feared situations. In such cases, an early responder is someone who has responded well to the initial sessions which is essentially restricted to the first or second components of CBT. In such cases, the skill sets or lack thereof and the needs for adequate recovery must be assumed to transfer from the early phases of treatment to performance on later phases. However, such transference may not be the case. Also, although a person may be a "slow starter" this does not mean that things won't "click" later leading to satisfactory results if treatment had continued rather than shifting to a new one because of futility. In many studies, early response is defined as a 20% or more reduction in symptoms from baseline by an a priori stated deadline, but if the crucial need for a patient is a topic considered later in the treatment protocol, then early non-response and a corresponding shift in treatment may deprive the person of that needed treatment. The identification of early responders versus non-responders often requires nuanced and complex decision making.

Finally, some experts argue against treatment shifts to either early responders (by ending treatment to them) or non-responders (by shifting treatments) because they see them as rushes to judgment. They argue one must give an intervention or treatment a chance to do what it was intended to do via the protocols established in prior large scale efficacy trials; that we undercut treatment viability by not letting it run its course.

Adaptive Designs

There are many forms of adaptive designs but a common one is known as SMART. This is an acronym for **Sequential Multiple Assignment Randomized Trial**. The key features of a SMART design are (a) participants move through multiple stages of treatment over time (sequential), (b) participants are randomized more than once at key decision points (multiple assignment) (c) participants receive tailored care. Re-randomization often depends on how well a patient responds to the first-stage treatment, which turns the design focus to one of responders versus non-responders. The idea is to use SMART designs to build optimized, adaptive interventions.

A variant of the SMART strategy is a **stepped-care design** also called a **response-conditional adaptive pathway design**. This approach involves only a single point of randomization typically at baseline. In some respects, it is a quasi-experimental design because for some facets of it, randomization is broken as I discuss shortyl. The core logic

goes something like this: In the real world, it is often the case that a treatment, such as a medication or cognitive behavior therapy CBT) is administered as a starting point. Some people respond well to the treatment but others do not. For responder status at midtreatment or some other key decision point, people who have responded well continue with the treatment protocol. The non-responders are either switched to a new treatment (e.g., a new medication because the first one was not working) or to the current treatment but now intensified or perhaps augmented by additional new strategies. The question then becomes (a) can people who started with the treatment but who were not showing a reasonable response ultimately achieve comparable levels of symptom relief as (b) those who received the treatment and responded well to it initially. An additional question is (c) does this treatment strategy as a whole (use of an adaptive approach for the two groups combined) exceed a control condition, such as treatment as usual.

Technically, the stepped care approach is not a SMART design because it lacks sequential randomization. However, it captures real-world clinical practice and logic. To be sure, scientific complications lurk in this design. Most notable is that dividing participants based on treatment response at midtreatment breaks the initial randomization moving forward; we are no longer comparing randomized groups when we compare the two intervention conditions at posttest. Responders likely are individuals who are more amenable to the initial therapy and for whatever reason are destined to do well by it. The non-responders may have treatment-resistant symptoms, they may have lower baseline motivation, and/or lower working memory capacity that affects response to an intervention. Of course, it is exactly because of these reasons that, say CBT, is augmented for non-responders, but the bottom line is that the two groups no longer necessarily have equal mean "baselines" when the midtreatment outcome status acts as a "baseline" and they are segregated into groups of responders and non-responders.

Some methodologists might be inclined to compare the posttest scores for the two intervention groups after statistically controlling for the "baseline" mid-treatment differences. This is not necessarily a wise strategy. The problem is that, by design, there is little overlap between the two intervention groups on the outcome at midtreatment and this property then undermines the ANCOVA strategy (see the discussion of covariate overlap in Chapter 15). If a responder at midtreatment for whatever reason happens to have the same score as a nonresponder at midtreatment, then these individuals can hardly be thought of as equivalent cases. Finally, minimal overlapping groups typically trigger Lord's paradox (Allison. 1990) which can then lead to uninterpretable mathematical artifacts. It are such dynamics that turn the stepped care design into a quasi-experimental design (but with exceptions that I note shortly).

Allison (1990) argues that when relatively stable characteristics or non-random

selection underlies baseline group differences it generally is better to work with raw change scores than with ANCOVA-like strategies that statistically control for the outcome baseline. I use Allison's approach in my numerical example for a stepped-care trial below. In the context of the current discussion, responders may show little change from midtreatment to the immediate posttest but non-responders may show non-trivial improvement between these time points because of the addition of the new or augmented intervention strategies. If the two intervention groups ultimately have comparable post-intervention scores, the implication is that the two groups took very different pathways to get to the same place. The responders likely took a steady, gradual path to their posttest scores whereas the augmented group stalled initially, experienced a treatment modification, and then underwent a catch-up phase.[20]

The presence of a control group also can aid interpretation in stepped-care designs. A control group allows one to make clear statements about the efficacy of the intervention protocol when the protocol is defined as the two intervention groups combined. In other words, the intervention protocol can be conceptualized as a strategy of leaving people alone who initially respond positively to the treatment but adopting a "Plan B" for people who do not respond positively. Comparing this overall intervention protocol against a control group preserves the initial baseline randomization. It answers the question "*is offering a stepped-care program better than treatment as usual or an a priori defined control group of interest?*

Numerical Example

I consider as a numerical example a hypothetical stepped care design in which responder versus non-responder status on the outcome Y is determined at midtreatment. Responders at midtreatment finished the standard protocol for the remaining sessions after midtreatment while the non-responders at midtreatment were provided with augmented intervention strategies. Non-responder status at midtreatment was determined not just from a person's Y score at that time but also other criteria such as their vigilance in completing homework tasks, formation of constructive alliances with the interventionist, and so on also were taken into account. The control group received no intervention but it could just as easily be a TAU. At baseline, all individuals were randomly assigned to either the intervention or control conditions. All individuals were assessed at

[20] In traditional RCTs, methodologists typically prefer ANCOVA-like strategies using the baseline as a covariate because the imbalances between groups reflect temporary fluctuations and, transitory-like constructs. The raw change approach, by contrast, is more appropriate when there is non-trivial, stable sources of self-selection into treatment groups per the discussion of Allison (1990).

midtreatment, at an immediate posttest, and at a one month follow-up. Baseline measures were obtained prior to the start of treatment for everyone.

Two presumed mediators of Y were targeted for change by the program, MA and MB, on the assumption that changes in them produce positive changes in Y. A baseline covariate, COV was measured and controlled as it was deemed to be relevant to increased statistical power in the analyses. Y, MA and MB were each scored on 0 to 100 metrics but to make matters a bit more interpretable were divided by 10 to place then on a 0 to 10 scale. They all had standard deviations near 10 on the original metrics and 1.0 on the transformed metrics. The intervention sought to increase scores on MA, MB and, Y. For this example, Y is pain relief, MA is use of effective pain coping strategies and MB is positive cognitive restructuring regarding pain experiences.

The first set of analyses I pursue preserves randomization and conducts a two group evaluation of (a) an adaptive treatment approach that combined responders and non-responders into a single group even though they were acted upon differently during treatment per se versus (b) the non-treated control group. After making conclusions using this randomized design, I turn to supplemental analyses to gain perspectives on how the responders and non-responders fared at key time points in the trial. This latter portion of the analysis is quasi-experimental in character.

I use piecewise SEM (LISEM) to address the standard three core questions for mediation analysis in RETs, i.e. (1) does the intervention have an overall meaningful effect on the outcome, (2) does the intervention have an overall meaningful effect on each of the presumed mediators, and (3) do the mediators meaningfully impact the outcome. I use LISEM because creating a single overarching FISEM model created too many estimation challenges, at least for the current example. For reasons discussed earlier and elaborated upon later, I decided not to use the baseline as a covariate. This sacrificed some statistical power but in the present case, doing so was not consequential.

Overall Effect of the Intervention on the Outcome

Before presenting the Mplus syntax about the overall effect of the intervention on the outcome, it is helpful to pre-shadow results by plotting the predicted covariate adjusted Y means for the three group design that separated responders and non-responders at midtreatment in conjunction with the control group and to also construct a plot for the two group design that collapsed responders and non-responders into a single intervention group. I used the *Temporal Line Plot* program on my website to do so. Figure 16.30 presents the three group plot. Time 0 is the baseline, time 1 is the midtreatment, time 2 is the immediate posttest, and time 3 is the one month follow-up. The control group shows a relatively flat mean profile across time. The non-responder group shows very little

change from baseline to midtreatment and then a sharp increase in Y (symptom relief) after the midtreatment when the treatment protocol was switched/adapted. The responder group shows good response at mid-treatment that persists in the later phases of the study.

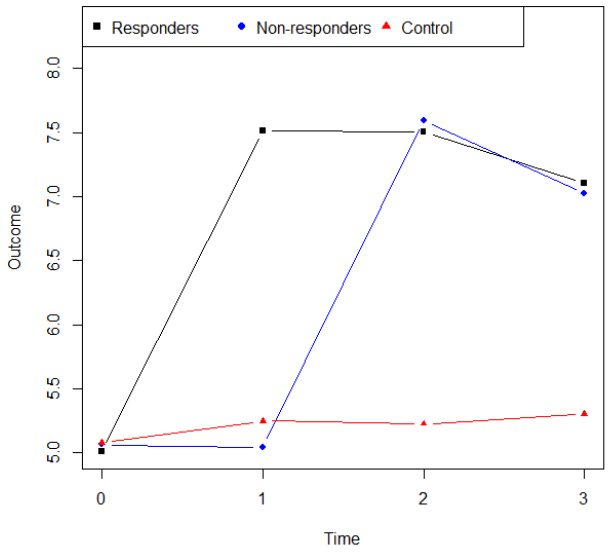


FIGURE 16.30 Three group adaptive design

Figure 16.31 presents the two group plot that collapses the responders and non-responders into a single group. The two group plot essentially averages the values for the responder and non-responder groups at a given time point relative to the three group plot.

My initial focus in this and most of the sections that follow is on the two group randomized design. As with previous examples in this chapter, there is not a single "total" effect in this study. Rather, there is a total effect at each post-baseline time point. My programming strategy is to use Mplus to generate predicted covariate adjusted outcome means for each cell of the 2X4 study design (treatment group by time) and then use the `MODEL CONSTRAINT` command to conduct substantively interesting contrasts on those means. For this first question, the focus is on two contrasts, (a) the intervention versus control group at the immediate posttest and (b) the intervention versus control group at the follow-up. The Mplus syntax performs the analysis simultaneously for both the outcomes and the mediators but you can conduct separate analyses on each variable type if desired. The program purposely is not embedded in a larger FISEM-like model that may be subject to complex specification errors. Rather, it fairly simply zeros in on isolating cell means and then performing substantive contrasts of interest among them.

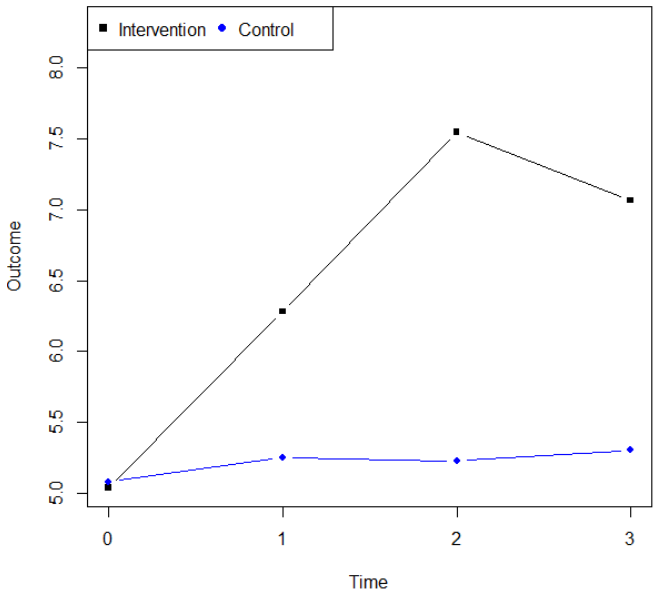


FIGURE 16.31 Two group adaptive design

Given a focus on the 2X4 study design, one might wonder why bother with SEM when, perhaps, a more traditional 2X4 ANOVA or ANCOVA might suffice using standard software. There are multiple reasons to prefer SEM modeling. First, traditional mixed ANOVAs of this type assume compound symmetry/sphericity (see Maxwell, Delaney & Kelley, 2017). The SEM approach, by contrast, estimates a saturated unstructured covariance matrix for the disturbances across time thereby bypassing compound symmetry assumptions. Second, SEM allows for the use of robust estimation, such as MLR or bootstrapping in Mplus. This eases normality assumptions. Indeed, it is possible in Mplus to use Bayesian estimation which offers many advantages relative to maximum likelihood more generally. Third, SEM enables use of full information maximum likelihood (FIML) for missing data, which is more elegant and often more efficient than listwise deletion that is typical in mixed ANOVAs (see Chapter 26). Fourth, SEM can accommodate latent variables with multiple indicators thereby making adjustments for measurement error, as needed.

Table 16.22 presents the Mplus syntax I used to generate the cell means and contrasts of interest for purposes of analyzing the total effect of the intervention on Y, and later for evaluating intervention effects on the mediators:

Table 16.22. Mplus Syntax for Two Group Adaptive Design

```
1.  TITLE: Piecewise SEM: Cell means followed by contrasts;
2.  DATA: FILE = adaptiveM.dat;
3.  DEFINE:
4.    center cov (grandmean) ;  !mean center covariate
5.  VARIABLE:
6.    NAMES ARE id group tx1_d tx2_d cov y0 ma0 mb0
7.    y1 ma1 mb1 y2 ma2 mb2 y3 ma3 mb3 treat;
8.  USEVARIABLES = treat cov y0 y1 y2 y3 ma0
9.    ma1 ma2 ma3 mb0 mb1 mb2 mb3;
10. ANALYSIS: ESTIMATOR = MLR ;
11. MODEL:
12.  !Make covariate adjustments
13.  y0 y1 y2 y3 ma0 ma1 ma2 ma3 mb0 mb1 mb2 mb3 ON cov;
14.  !Regress wave measurements on group dummy var
15.  !Outcome Y
16.  [y0 y1 y2 y3] (cy0 cy1 cy2 cy3);  !estimate intercepts
17.  y0 ON treat (dy1_0 );
18.  y1 ON treat (dy1_1 );
19.  y2 ON treat (dy1_2) ;
20.  y3 ON treat (dy1_3) ;
21.  !Mediator MA
22.  [ma0 ma1 ma2 ma3] (ca0 ca1 ca2 ca3);  !estimate intercepts
23.  ma0 ON treat (da1_0 );
24.  ma1 ON treat (da1_1 );
25.  ma2 ON treat (da1_2 );
26.  ma3 ON treat (da1_3 );
27.   !Mediator MB
28.  [mb0 mb1 mb2 mb3] (cb0 cb1 cb2 cb3);  !estimate intercepts
29.  mb0 ON treat (db1_0 );
30.  mb1 ON treat (db1_1 );
31.  mb2 ON treat (db1_2 );
32.  mb3 ON treat (db1_3 );
33.  !correlate vars across time
34.  y0 WITH y1 y2 y3; y1 WITH y2 y3; y2 WITH y3;
35.  ma0 WITH ma1 ma2 ma3; ma1 WITH ma2 ma3; ma2 WITH ma3;
36.  mb0 WITH mb1 mb2 mb3; mb1 WITH mb2 mb3; mb2 WITH mb3;
37. MODEL CONSTRAINT:
38.  !Define cell means for Y
39. NEW(y_tx0_0  y_tx0_1  y_tx0_2  y_tx0_3
40.  y_tx1_0  y_tx1_1  y_tx1_2  y_tx1_3 );
41.  y_tx0_0 = cy0;
42.  y_tx0_1 = cy1;
43.  y_tx0_2 = cy2;
44.  y_tx0_3 = cy3;
45.  y_tx1_0 = cy0 + dy1_0;
46.  y_tx1_1 = cy1 + dy1_1;
47.  y_tx1_2 = cy2 + dy1_2;
```

```
48.  y_tx1_3 = cy3 + dy1_3;
49.  !Define cell means for MA
50. NEW(ma_tx0_0 ma_tx0_1 ma_tx0_2 ma_tx0_3
51.  ma_tx1_0 ma_tx1_1 ma_tx1_2 ma_tx1_3);
52.  ma_tx0_0 = ca0;
53.  ma_tx0_1 = ca1;
54.  ma_tx0_2 = ca2;
55.  ma_tx0_3 = ca3;
56.  ma_tx1_0 = ca0 + da1_0;
57.  ma_tx1_1 = ca1 + da1_1;
58.  ma_tx1_2 = ca2 + da1_2;
59.  ma_tx1_3 = ca3 + da1_3;
60.  !Define cell means for MB
61. NEW(mb_tx0_0 mb_tx0_1 mb_tx0_2 mb_tx0_3
62.  mb_tx1_0 mb_tx1_1 mb_tx1_2 mb_tx1_3);
63.  mb_tx0_0 = cb0;
64.  mb_tx0_1 = cb1;
65.  mb_tx0_2 = cb2;
66.  mb_tx0_3 = cb3;
67.  mb_tx1_0 = cb0 + db1_0;
68.  mb_tx1_1 = cb1 + db1_1;
69.  mb_tx1_2 = cb2 + db1_2;
70.  mb_tx1_3 = cb3 + db1_3;
71.  !Pairwise contrasts post-baseline for Y, MA and MB
72. NEW(y_d1_t1  y_d1_t2  y_d1_t3
73.  ma_d1_t1  ma_d1_t2  ma_d1_t3
74.  mb_d1_t1  mb_d1_t2  mb_d1_t3 );
75.  !Y Contrasts
76.  y_d1_t1 = y_tx1_1 - y_tx0_1; !Tx vs control mid (t1)
77.  y_d1_t2 = y_tx1_2 - y_tx0_2; !Tx vs control post (t2)
78.  y_d1_t3 = y_tx1_3 - y_tx0_3; !Tx vs control fu (t3)
79.  !MA Contrasts
80.  ma_d1_t1 = ma_tx1_1 - ma_tx0_1; !Tx vs control mid (t1)
81.  ma_d1_t2 = ma_tx1_2 - ma_tx0_2; !Tx vs control post (t2)
82.  ma_d1_t3 = ma_tx1_3 - ma_tx0_3; !Tx vs control fu (t3)
83.  !MB Contrasts
84.  mb_d1_t1 = mb_tx1_1 - mb_tx0_1; !Tx vs control mid (t1)
85.  mb_d1_t2 = mb_tx1_2 - mb_tx0_2; !Tx vs control post (t2)
86.  mb_d1_t3 = mb_tx1_3 - mb_tx0_3; !Tx vs control fu (t3)
87.  !Example decay analysis
88. NEW (decay32) ;
89.  decay32=(y_tx1_3 - y_tx1_2 ) - (y_tx0_3 -  y_tx0_2) ;
90. OUTPUT: SAMP STDYX RESUDAL CINTERVAL TECH4;
```

Coupled with the comment lines, most of the syntax should be familiar based on prior examples we have considered in this chapter. Lines 3 and 4 grand mean center the covariate so that its overall mean value is zero. This removes the need to specify its formal mean value when generating the cell means in the later Mplus syntax. Line 13

regresses each of the relevant endogenous variables onto the covariate to adjust for the effects of the covariate. Each variable listed to the left of `ON` are separately regressed onto the variable to the right of `ON`. Lines 17 to 20 regress each of the Y variables onto the treatment dummy variable (0 = control, 1 = intervention) and assigns labels in parentheses to each coefficient. Note that even though I specified `cov` and `treat` as predictors of the Y on separate lines of syntax, Mplus merges them to be joint predictors of each Y in a multivariate sense. Line 16 estimates and labels each of the intercepts in these equations. Lines 21 to 26 repeat this process for the MA mediator and Lines 27 to 32 do so for the MB mediator. Lines 34 to 36 introduce correlated disturbances for the respective Y, MA and MB, which is consistent with their repeated measure character.

The `MODEL CONSTRAINT` commands in Lines 38 to 70 define the 2X4 cell means for the study design. The operations make use of the prior labels in the syntax for the respective regression intercepts and path coefficients and invoke principles previously discussed in this chapter (see also Appendix A). Lines 78 to 87 define pairwise contrasts between the means comparing the intervention and control conditions at midtreatment, at the immediate posttest, and at the follow-up for Y, MA, and MB, respectively. Lines 88 and 89 conduct an illustrative analysis of intervention decay that I describe later.

The model is just identified so matters of model fit are moot. Here are the results for the cell means followed by the contrasts for the outcome at the midtreatment, immediate posttest, and follow up (I add notes in **red** to help interpretation of the labels in the three rows beneath the notes) :

```
MODEL RESULTS

                                                    Two-Tailed
                  Estimate       S.E.  Est./S.E.    P-Value
New/Additional Parameters
    Control group (TX0) mid, post, fu means
    Y_TX0_1            5.247      0.091     57.897      0.000
    Y_TX0_2            5.226      0.091     57.593      0.000
    Y_TX0_3            5.302      0.106     50.048      0.000
    Tx group (TX1) mid, post, fu means
    Y_TX1_1            6.281      0.095     65.825      0.000
    Y_TX1_2            7.548      0.066    115.081      0.000
    Y_TX1_3            7.065      0.067    106.162      0.000
    Tx-Ctrl mid mean diff, post mean diff, fu mean diff
    Y_D1_T1            1.034      0.131      7.873      0.000
    Y_D1_T2            2.322      0.112     20.732      0.000
    Y_D1_T3            1.763      0.125     14.100      0.000
```

At the immediate posttest, the predicted mean for the intervention group was 7.55 ±0.13 and for the control group it was 5.23 ±0.18. The difference between the two predicted means was 2.32 ±0.22 (CR = 20.73, $p < 0.05$) or about 23 points on the original 0 to 100 metric. At the follow-up, the predicted mean for the intervention group was 7.07 ±0.13 and for the control group it was 5.30 ±0.21. The difference between the two predicted means was 1.76 ±0.25 (CR = 14.10, $p < 0.05$) or about 18 points on the original metric. Suppose the program staff and research team decided *a priori* that a mean difference of 10 points on the original metric of Y or 1.0 on the transformed metric was the standard for a meaningful population effect. The 95% confidence interval for the immediate posttest group difference on the transformed metric was 2.10 to 2.54 and at the follow up it was 1.51 to 2.01. In both cases, the lower limit of the intervals exceed the meaningfulness standard so both effects are declared meaningful.

Because multiple contrasts were performed, some researchers apply a Holm modified Bonferroni test or a False Discovery Rate (FDR) adjustment to adjust for test multiplicity, although as discussed in Chapter 6, some methodologists argue against such corrections. The effects remained statistically significant with or without the corrections so the point is somewhat moot. You can apply the corrections using the *FDR p values* and *Holm p values* programs on the *Programs* tab of my website.

Although I report the mid-treatment results above, when evaluating total treatment effects one typically does not evaluate the operative dynamics with midtreatment data. I consider such dynamics in my supplemental analyses later.

Effect of the Intervention on the Mediators

I evaluate the effect of the intervention on the two mediators using the same strategy as that for the outcome. I omit the mean plots in the interest of space. Here are the relevant predicted mean values and contrasts:

```
MODEL RESULTS

                                                    Two-Tailed
                  Estimate       S.E.  Est./S.E.    P-Value
New/Additional Parameters
    Control group (TX0) mid, post, fu means
    MA_TX0_1           4.913      0.081     60.378      0.000
    MA_TX0_2           4.891      0.081     60.333      0.000
    MA_TX0_3           4.918      0.083     59.052      0.000
    Tx group (TX1) mid, post, fu means
    MA_TX1_1           6.156      0.094     65.560      0.000
    MA_TX1_2           7.465      0.061    121.419      0.000
    MA_TX1_3           7.415      0.058    126.812      0.000
```

```
    Tx-Ctrl mid mean diff, post mean diff, fu mean diff
    MA_D1_T1           1.243      0.124     10.023      0.000
    MA_D1_T2           2.575      0.102     25.325      0.000
    MA_D1_T3           2.497      0.102     24.572      0.000
```

At the immediate posttest, the predicted MA mean for the intervention group was 7.46 ±0.12 and for the control group it was 4.89 ±0.16. The difference between the two predicted means was 2.58 ±0.21 (CR = 25.32, $p < 0.05$). At the follow-up, the predicted MA mean for the intervention group was 7.41 ±0.12 and for the control group it was 4.92 ±0.17. The difference between the two predicted means was 2.50 ±0.20 (CR = 24.57, $p < 0.05$). Suppose the program staff and research team decided *a priori* that a mean difference of 10 points on the original 0 to 100 metric of MA, or 1.0 on the transformed metric, was the standard for a meaningful population effect. The 95% confidence interval for the immediate posttest group difference on the transformed metric was 2.37 to 2.77 and at the follow up it was 2.30 to 2.70. In both cases, the lower limit of the intervals exceeds the meaningfulness standard so both effects are declared meaningful.

Here are the relevant predicted mean values and contrasts for MB:

```
                                                    Two-Tailed
                   Estimate       S.E.  Est./S.E.    P-Value
New/Additional Parameters
    Control group (TX0) mid, post, fu means
    MB_TX0_1           5.106      0.080     63.596      0.000
    MB_TX0_2           5.037      0.074     68.069      0.000
    MB_TX0_3           5.140      0.076     67.488      0.000
    Tx group (TX1) mid, post, fu means
    MB_TX1_1           6.034      0.058    103.938      0.000
    MB_TX1_2           5.987      0.057    105.575      0.000
    MB_TX1_3           6.055      0.057    105.506      0.000
    Tx-Ctrl mid mean diff, post mean diff, fu mean diff
    MB_D1_T1           0.928      0.099      9.351      0.000
    MB_D1_T2           0.950      0.093     10.198      0.000
    MB_D1_T3           0.915      0.095      9.584      0.000
```

At the immediate posttest, the predicted MB mean for the intervention group was 5.99 ±0.12 and for the control group it was 5.04 ±0.15. The difference between the two predicted means was 0.95 ±0.19 (CR = 10.20, $p < 0.05$) or about 9 units on the untransformed 0 to 100 MB metric. At the follow-up, the predicted MB mean for the intervention group was 6.06 ±0.11 and for the control group it was 5.14 ±0.15. The difference between the two predicted MB means was 0.92 ±0.19 (CR = 9.58, $p < 0.05$). Suppose the program staff ad research team decided *a priori* that a mean difference of 10 points on the original 0 to 100 metric of MB or 1.0 on the transformed metric was the

standard for a meaningful population effect. The 95% confidence interval for the immediate posttest group difference on the transformed metric was 0.76 to 1.14 and at the follow up it was 0.73 to 1.11. In both cases, the confidence intervals overlapped the meaningfulness standard indicating we cannot confidently conclude the effects are meaningful. Based on the significance tests, we can conclude the effects on MB are non-zero by virtue of the null hypothesis tests but after taking into account sampling error, we cannot conclude the effects are meaningful.

Effect of the Mediators on the Outcome

Given the presence of four time points (baseline, midtreatment, immediate posttest and follow-up) I decided to evaluate the effect of the mediators on the outcome using Allison et al.'s (2017) SEM-fixed effect approach discussed earlier in the chapter. This strategy is within-person based and has the advantage of controlling for all between-person confounds of the M→Y links. As such, it also controls for the direct effect of the treatment condition on the outcome as well as the effects of the covariate on the outcome. Table 16.23 presents the relevant syntax.

Table 16.23. Fixed Effect Model for Two Group Adaptive Design

```
1.  TITLE: Piecewise SEM: Allison SEM Fixed Effects;
2.  DATA: FILE = adaptiveM.dat ;
3.  VARIABLE:
5.  VARIABLE:
6.    NAMES ARE id group tx1_d tx2_d cov y0 ma0 mb0
7.    y1 ma1 mb1 y2 ma2 mb2 y3 ma3 mb3 treat;
8.  USEVARIABLES = y0 y1 y2 y3 ma0 ma1 ma2 ma3 mb0 mb1 mb2 mb3
9.    mb0 mb1 mb2 mb3 ;
10. ANALYSIS: ESTIMATOR = MLR ;
11. MODEL:
12.   alpha BY y0@1 y1@1 y2@1 y3@1; !define fixed effect
13.  !Set causal paths with equality constraints
14,   y0 ON ma0 mb0 (w_ma w_mb);
15.   y1 ON ma1 mb1 (w_ma w_mb);
16.   y2 ON ma2 mb2 (w_ma w_mb);
17.   y3 ON ma3 mb3 (w_ma w_mb);
18.   y0 ;   y1 y2 y3 (evar); !Constrain resid vars to equal except y0
19.   alpha WITH ma0-ma3 mb0-mb3; !Cor latent var with preds
20. OUTPUT: SAMP MOD(4) STDYX RESIDUAL CINTERVAL TECH4;
```

With one exception, I do not comment on the syntax because it follows directly from the syntax in Table 16.5. Line 18 estimates the disturbance variances of the Y while imposing an equality constraint on them by virtue of the use of the common label `evar`.

Standard practice is to apply the equality constraint to all Y but in the current case I relaxed the constraint but only for the baseline Y. I did so because in an initial model analysis, the modification indices indicated that this variance needed to be estimated separately. Doing so also makes conceptual sense, so I removed the label from `y0` to relax the constraint during model estimation.

The model fit for the SEM-Based fixed effects model was reasonable although the chi square test of perfect model fit in the population was statistically significant (chi square = 97.88 with 29 degrees of freedom, $p < 0.05$). In the current case, the sample size was fairly large N = 450, which tends to magnify small disparities between predicted and observed covariances. None of the modification indices suggested meaningful nor theoretically coherent ill fit nor was this the case for the analysis of standardized residuals. Also, since I generated the data for the population in a way that conformed to the model, I know for a fact that the current result is chance generated (which is not something one normally would know). The other fit indices all suggested reasonable model fit. The RMSEA was 0.073 with an upper limit for the 90% confidence interval of 0.082. The CFI was 0.98 and the standardized RMR was 0.035. Overall, I move forward with the analysis but I also perform additional sensitivity checksbelow.

Here are the results for the within-person regression, shown here for the immediate posttest but that are equal the coefficients at every time point because of the imposed equality constraints on the path coefficients in an SEM-based fixed effects analysis:

```
MODEL RESULTS

                                                    Two-Tailed
                  Estimate       S.E.  Est./S.E.    P-Value

 Y2       ON
    MA2               0.742      0.012     61.485      0.000
    MB2               0.126      0.020      6.222      0.000
```

The coefficient for the effect of the mediator MA on Y was 0.74 (±0.02, CR = 61.48, $p < 0.05$) and for MB it was 0.13 (±0.04, CR = 6.22, $p < 0.05$). For MA, for every one unit that MA increases across time for people, Y is predicted to increase by 0.74 using the transformed metrics. Suppose prior to the study it was decided that a population coefficient of 0.10 or greater on the transformed metric is meaningful. The 95% confidence interval for the MA coefficient was 0.72 to 0.76. Because the lower limit of the interval for the coefficient is greater than the standard, the MA→Y effect is declared meaningful. For MB, suppose prior to the study it was decided that a population coefficient of 0.10 or greater on the transformed metric also was a reasonable

meaningfulness standard. The 95% confidence interval for the MB coefficient was 0.09 to 0.16. Because the confidence interval overlaps the standard, we cannot conclude the effect is meaningful. Given the statistical significance of the coefficient ($p < 0.05$) we can confidentially conclude the population coefficient is not zero; and that the sample value of the coefficient (0.13) is suggestive of meaningfulness. However, after taking into account sampling error, we cannot say with confidence that the magnitude of the coefficient is meaningful.

For sensitivity purposes, I also performed this analysis using the program called *GLM panel regression* on the Programs tab of my website rather than Allison's SEM approach. It is less flexible than the SEM strategy, uses slightly different estimation strategies, and requires long format data. Here is the output for it:

	Est.	S.E.	t val.	d.f.	p
(Intercept)	6.060	0.056	107.895	449.000	0.000
ma	0.806	0.013	61.182	1348.000	0.000
mb	0.143	0.024	6.029	1348.000	0.000

The results for the within-person analyses are similar to those for Allison's SEM approach.

In some cases, researchers evaluate between-person coefficients for the effects of MA and MB on Y using more traditional regression strategies. In the current case, the traditional approach would regress Y at the immediate posttest onto MA, MB, TREAT and COV at the immediate posttest using standard statistical software or Mplus. This turns out to be problematic with the current data because of a dynamic I identified in Chapter 9. Specifically, the effect of the treatment condition on MA was so strong that these two predictors suffered from a high degree of multicollinearity ($r = 0.76$) which undermines the analysis and makes the regression highly unstable. There are no satisfactory ways of dealing with this matter in the current case; rather, the data are such that it simply is not possible to reliably isolate a between-person effect for MA and MB on Y while simultaneously controlling for the direct effect of T→Y. Because I already have reasonable estimates of the effect of MA and MB on Y independent of TREAT from the SEM-based fixed effects analysis and the GLM fixed effect analysis, I decide to be content with those estimates and move forward with my conclusions accordingly. The bottom line is that there is reasonable support for the effects of the mediators on the outcome.

Supplemental Analyses

To gain perspectives on the possible impact of making an adaptive change in treatment for non-responders, I repeated the two-group analysis that isolated cell means for the 2X4 design but now I split the intervention group into two segments, responders by midtreatment versus non-responders by midtreatment to yield a 3X4 design. The first factor consisted of the three treatment conditions (treated initial responders versus treated initial non-responders versus control) and the second factor was time (baseline, midtreatment, immediate posttest and follow up). Table 16.24 presents the Mplus syntax for the analysis.

Table 16.24. Mplus Syntax for Three Group Adaptive Design

```
1.   TITLE: Piecewise SEM: Cell means followed by contrasts;
2.    DATA: FILE = adaptiveM.dat;
3.    DEFINE:
4.      center cov (grandmean) ;  !mean center covariate
5.    VARIABLE:
6.      NAMES ARE id group tx1_d tx2_d cov y0 ma0 mb0
7.      y1 ma1 mb1 y2 ma2 mb2 y3 ma3 mb3 treat;
8.      USEVARIABLES = tx1_d tx2_d cov y0 y1 y2 y3 ma0
9.      ma1 ma2 ma3 mb0 mb1 mb2 mb3;
10.   ANALYSIS: ESTIMATOR = MLR ;
11.   MODEL:
12.    !Make covariate adjustments
13.    y0 y1 y2 y3 ma0 ma1 ma2 ma3 mb0 mb1 mb2 mb3 ON cov;
14.    !Regress wave measurements on group dummy variables
15.    !Outcome Y
16.    [y0 y1 y2 y3] (cy0 cy1 cy2 cy3);  !estimate intercepts
17.    y0 ON tx1_d tx2_d (dy1_0 dy2_0);
18.    y1 ON tx1_d tx2_d (dy1_1 dy2_1);
19.    y2 ON tx1_d tx2_d (dy1_2 dy2_2);
20.    y3 ON tx1_d tx2_d (dy1_3 dy2_3);
21.    !Mediator MA
22.    [ma0 ma1 ma2 ma3] (ca0 ca1 ca2 ca3);  !estimate intercepts
23.    ma0 ON tx1_d tx2_d (da1_0 da2_0);
24.    ma1 ON tx1_d tx2_d (da1_1 da2_1);
25.    ma2 ON tx1_d tx2_d (da1_2 da2_2);
26.    ma3 ON tx1_d tx2_d (da1_3 da2_3);
27.    !Mediator MB
28.    [mb0 mb1 mb2 mb3] (cb0 cb1 cb2 cb3);  !estimate intercepts
29.    mb0 ON tx1_d tx2_d (db1_0 db2_0);
30.    mb1 ON tx1_d tx2_d (db1_1 db2_1);
31.    mb2 ON tx1_d tx2_d (db1_2 db2_2);
32.    mb3 ON tx1_d tx2_d (db1_3 db2_3);
33.    !correlate vars across time
34.    y0 WITH y1 y2 y3; y1 WITH y2 y3; y2 WITH y3;
35.    ma0 WITH ma1 ma2 ma3; ma1 WITH ma2 ma3; ma2 WITH ma3;
```

```
 mb0 WITH mb1 mb2 mb3; mb1 WITH mb2 mb3; mb2 WITH mb3;
MODEL CONSTRAINT:
!Define cell means for Y
NEW(y_tx0_0  y_tx0_1  y_tx0_2  y_tx0_3
 y_tx1_0  y_tx1_1  y_tx1_2  y_tx1_3
 y_tx2_0  y_tx2_1  y_tx2_2  y_tx2_3);
 y_tx0_0 = cy0;
 y_tx0_1 = cy1;
 y_tx0_2 = cy2;
 y_tx0_3 = cy3;
 y_tx1_0 = cy0 + dy1_0;
 y_tx1_1 = cy1 + dy1_1;
 y_tx1_2 = cy2 + dy1_2;
 y_tx1_3 = cy3 + dy1_3;
 y_tx2_0 = cy0 + dy2_0;
 y_tx2_1 = cy1 + dy2_1;
 y_tx2_2 = cy2 + dy2_2;
 y_tx2_3 = cy3 + dy2_3;
NEW(ma_tx0_0 ma_tx0_1 ma_tx0_2 ma_tx0_3
 ma_tx1_0 ma_tx1_1 ma_tx1_2 ma_tx1_3
 ma_tx2_0 ma_tx2_1 ma_tx2_2 ma_tx2_3);
 ma_tx0_0 = ca0;
 ma_tx0_1 = ca1;
 ma_tx0_2 = ca2;
 ma_tx0_3 = ca3;
 ma_tx1_0 = ca0 + da1_0;
 ma_tx1_1 = ca1 + da1_1;
 ma_tx1_2 = ca2 + da1_2;
 ma_tx1_3 = ca3 + da1_3;
 ma_tx2_0 = ca0 + da2_0;
 ma_tx2_1 = ca1 + da2_1;
 ma_tx2_2 = ca2 + da2_2;
 ma_tx2_3 = ca3 + da2_3;
NEW(mb_tx0_0 mb_tx0_1 mb_tx0_2 mb_tx0_3
 mb_tx1_0 mb_tx1_1 mb_tx1_2 mb_tx1_3
 mb_tx2_0 mb_tx2_1 mb_tx2_2 mb_tx2_3);
 mb_tx0_0 = cb0;
 mb_tx0_1 = cb1;
 mb_tx0_2 = cb2;
 mb_tx0_3 = cb3;
 mb_tx1_0 = cb0 + db1_0;
 mb_tx1_1 = cb1 + db1_1;
 mb_tx1_2 = cb2 + db1_2;
 mb_tx1_3 = cb3 + db1_3;
 mb_tx2_0 = cb0 + db2_0;
 mb_tx2_1 = cb1 + db2_1;
 mb_tx2_2 = cb2 + db2_2;
 mb_tx2_3 = cb3 + db2_3;
!Pairwise contrasts post-baseline for Y, MA and MB
NEW(y_d1_t1   y_d2_t1   y_d3_t1
 y_d1_t2   y_d2_t2   y_d3_t2
 y_d1_t3   y_d2_t3   y_d3_t3
```

```
88.   ma_d1_t1  ma_d2_t1  ma_d3_t1
89.   ma_d1_t2  ma_d2_t2  ma_d3_t2
90.   ma_d1_t3  ma_d2_t3  ma_d3_t3
91.   mb_d1_t1  mb_d2_t1  mb_d3_t1
92.   mb_d1_t2  mb_d2_t2  mb_d3_t2
93.   mb_d1_t3  mb_d2_t3  mb_d3_t3);
94.   !Y contrasts midtreatment (T1)
95.   y_d1_t1 = y_tx1_1 - y_tx0_1; ! Tx1 vs Control
96.   y_d2_t1 = y_tx2_1 - y_tx0_1; ! Tx2 vs Control
97.   y_d3_t1 = y_tx2_1 - y_tx1_1; ! Tx2 vs Tx1
98.   !Y contrasts immediate posttest (T2)
99.   y_d1_t2 = y_tx1_2 - y_tx0_2; ! Tx1 vs Control
100.  y_d2_t2 = y_tx2_2 - y_tx0_2; ! Tx2 vs Control
101.  y_d3_t2 = y_tx2_2 - y_tx1_2; ! Tx2 vs Tx1
102.  !Y contrasts follow up (T3)
103.  y_d1_t3 = y_tx1_3 - y_tx0_3; ! Tx1 vs Control
104.  y_d2_t3 = y_tx2_3 - y_tx0_3; ! Tx2 vs Control
105.  y_d3_t3 = y_tx2_3 - y_tx1_3; ! Tx2 vs Tx1
106.  !MA contrasts midtreatment (T1)
107.  ma_d1_t1 = ma_tx1_1 - ma_tx0_1;
108.   ma_d2_t1 = ma_tx2_1 - ma_tx0_1;
109.   ma_d3_t1 = ma_tx2_1 - ma_tx1_1;
110.   !MA contrasts immediate posttest (T2)
111.   ma_d1_t2 = ma_tx1_2 - ma_tx0_2;
112.   ma_d2_t2 = ma_tx2_2 - ma_tx0_2;
113.   ma_d3_t2 = ma_tx2_2 - ma_tx1_2;
114.   !MA contrasts followup (T3)
115.   ma_d1_t3 = ma_tx1_3 - ma_tx0_3;
116.   ma_d2_t3 = ma_tx2_3 - ma_tx0_3;
117.   ma_d3_t3 = ma_tx2_3 - ma_tx1_3;
118.   !MB contrasts midtreatment (T1)
119.   mb_d1_t1 = mb_tx1_1 - mb_tx0_1;
120.   mb_d2_t1 = mb_tx2_1 - mb_tx0_1;
121.  mb_d3_t1 = mb_tx2_1 - mb_tx1_1;
122.  !MB contrasts immediate posttest(T2)
123.  mb_d1_t2 = mb_tx1_2 - mb_tx0_2;
124.  mb_d2_t2 = mb_tx2_2 - mb_tx0_2;
125.  mb_d3_t2 = mb_tx2_2 - mb_tx1_2;
126.  !MB contrasts follow up (T3)
127.  mb_d1_t3 = mb_tx1_3 - mb_tx0_3;
128.  mb_d2_t3 = mb_tx2_3 - mb_tx0_3;
129.  mb_d3_t3 = mb_tx2_3 - mb_tx1_3;
130.  !Effect of switch for non-responders
131. NEW (nonr2_1 ctrl2_1 switch) ;
132.  nonr2_1 = (y_tx1_2 - y_tx1_1)  ;
133.  ctrl2_1 = (y_tx0_2 - y_tx0_1) ;
134.  switch = nonr2_1 - ctrl2_1 ;
135. OUTPUT: SAMP STDYX MOD(4) CINTERVAL TECH4;
```

No commentary on the syntax is required because it is a straightforward extension of the syntax in Table 16.22 but with additional contrasts in Lines 131 to 134 which I

explain shortly. The model is just-identified so issues of model fit are moot. I segregate here a subset of the predicted cell means for Y from the analysis to help organize further discussion of them:

Group	Midtreat	Immed Post
Responders	7.515	7.505
Non-responders	5.044	7.591
Controls	5.249	5.226

As expected, the non-responders exhibit Y means comparable to the control group at midtreatment and substantially lower than the Y means for the responders at midtreatment. Two questions are of interest. The first question asks whether the mean change in Y from the midtreatment to the immediate posttest for the non-responders (from 5.044 to 7.591) meaningfully exceeds the change in the corresponding means for the control group (from 5.249 to 5.226). The second question asks whether the level of relief at the immediate posttest for the initial non-responders (7.591) has rebounded to the extent it reaches comparable levels to the responders at the immediate posttest (7.505).

When groups differ non-trivially in their functional "baseline" values, which are the midtreatment values here, Allison (1990) recommends focusing on mean changes rather than "baseline" covariate adjusted mean differences to avoid complications due to Lord's paradox. In the `MODEL CONSTRAINT` commands on Lines 131-134 I included contrasts that compared the classic difference in mean differences (DiD in the economics literature) for the change for non-responders compared to the change for controls. Here are the relevant results:

```
                                                    Two-Tailed
                  Estimate       S.E.  Est./S.E.    P-Value

New/Additional Parameters
    NONR2_1          2.547      0.067     38.024      0.000
    CTRL2_1         -0.024      0.071     -0.333      0.739
    SWITCH           2.571      0.098     26.342      0.000
```

The estimated Y mean change for non-responders calculated by subtracting the midtreatment mean from the immediate posttest mean was 7.59 – 5.04 = 2.55 ± 0.13 (CR = 38.02, $p < 0.05$). For the control group, the mean change was 5.27 – 5.259 = -0.02 ±0.14 (CR = 0.33, *ns*). The difference between these differences was 2.57 ±0.20 (CR = 26.34, $p < 0.05$) suggesting the switch from standard treatment to a modified treatment improved Y for the non-responders (using a meaningfulness standard of 1.0).

Here is the contrast from Line 101 that addressed the second question by subtracting the estimated mean at the immediate posttest for the non-responders (7.591) from the estimated mean at the immediate posttest for the responders (7.505):

```
                                                    Two-Tailed
                    Estimate       S.E.  Est./S.E.    P-Value

New/Additional Parameters
   Y_D3_T2           -0.086      0.131     -0.657      0.511
```

The mean difference was -0.09 ±0.26 (CR = 0.66, *ns*) suggesting a trivial difference. If desired, I can do a formal test of functional equivalence between the means as outlined in Chapters 2 and 10. Suppose I define the latitude of no effect as a population mean difference between the values of -1.00 and 1.00. The 95% confidence interval for the observed above mean difference is -0.35 to 0.18. Because the confidence interval is completely contained within the latitude of no effect, I declare the two means as being functionally equivalent.

In sum, there appears to be empirical support that the treatment protocol switch for early non-responders made a meaningful change in Y and that it had the effect of bringing them up to comparable levels of Y for early responders who were not switched from the initial treatment. Having said that, there are factors that complicate this interpretation.

One issue is that of regression to the mean as an artifact relative to the estimated intervention effect for non-responders from midtreatment to the immediate posttest. The general logic goes something like this: At the midterm assessment, some participants could have been placed in the non-responder group not because they deserved to be there but instead because they just happened to have a particularly bad day or two in terms of experienced pain that artificially pushed their midtreatment Y score downward. Or perhaps random noise associated with measurement error artificially pushed their observed Y score at midtreatment downward relative to their true scores. At the later immediate posttest, these sources of random noise likely would not occur because they are random not systematic. The result would be an artificial increase in the observed Y measure for these individuals, making the treatment switch appear more influential than it is. This reflects the phenomena of regression to the mean described in Chapter 4 and is particularly problematic when extreme groups are studied.

The same dynamics would hold for individuals in the responder group. At the midterm assessment, some participants could have been placed in the responder group not because they deserved to be there but instead because they just happened to have a particularly good day or two in terms of experienced pain. Or perhaps the random noise

associated with measurement error artificially pushed their observed Y score upward relative to their true score. At the immediate posttest, these sources of random noise likely would not occur, the result of which would an artificial decrease in the observed Y measure, making the treatment appear less influential artificially.

Usually, regression to the mean operates to the same extent for people selected into the two "extreme" groups, i.e., the amount of artificial decay in the responders will equal the amount of artificial increase in the non-responders. Note that in the current data, there was almost no change for the responders from mid-treatment to the immediate posttest but that for the non-responders, the improvement in scores was considerable. This suggests that if regression to the mean is operating, it is doing so in different ways in the two groups.

One challenge with regression to the mean explanations is deciding what population mean the respective groups are regressing toward. In the present case, one view is that the true population combined mean for responders and non-responders at midtreatment is a mean midway between their respective midtreatment mean scores which in the current case would be a value near 6.25. Another view is that non-responders are a distinct sub-population from responders with their own true underlying pain relief/severity mean and that regression to the mean dynamics operate relative to their own mean, not the combined mean of responders and non-responders. Similarly so for the responder group.

Mee and Chua (1991) developed a method for dependent pre-post comparisons that estimate the difference between the two means after adjusting for regression to the mean. The method is available in the program *Regression to mean* on the Programs tab of my website. The progam requires the user to "take a stand" and specify the operative population mean that the two groups regress toward. I used a population mean of 6.25. The unadjusted mean difference from midtreatment to the immediate posttest for non-responders was 2.55 whereas the adjusted mean difference using the Mee and Chua correction was 2.19, which remained statistically significant after adjustment for regression to the mean ($t = 22.97$, $p < 0.05$). Regression to the mean accounts for about $2.55 - 2.19 = 0.36$ of the 2.55 increase from midtreatment to the immediate posttest. Overall, it seems that the switch of the non-responders to an augmented treatment was beneficial, although one should make such a conclusion with humility.

The best way to deal with regression to mean and other potential confounds that the broken randomization creates is to use a SMART design by randomly assigning the non-responders at midtreatment to either (a) a non-responder group that then receives the new/augmented intervention or to (b) an appropriately defined control group (e.g. a group that continues with the original treatment despite poor performance). However, this

approach becomes more costly from a practical standpoint to the point and it may not be feasible. The choice of an appropriate control condition also can be controversial.

I could pursue comparable analyses for MA and MB to gain additional insights into the above dynamics but in the interest of space, I leave that as an exercise for you. The analyses follow the same roadmap as that outlined above for the outcome. The raw data I analyzed are provided on the Resources tab of my website under Chapter 16.

Omnibus Mediation and Concluding Comments for the Adaptive Example

Although they are of lower priority for me, it is straightforward to conduct null hypothesis tests of omnibus mediation using the joint significance test. Both MA and MB exhibited statistically significant ($p < 0.05$) T→M and M→Y links suggesting the presence of some degree of mediation. However, the intervention failed to have a meaningful effect on MB nor did MB yield a meaningful effect on Y. Perhaps program designers need to more closely examine their intervention strategies for addressing MB and whether the time addressing MB is well spent.

The strategy I used for analyzing core parts of the data relied on contrast analyses of the predicted means. The main Mplus program generated the cell means for the study design and the `MODEL CONSTRAINTS` commands were used to conduct the substantive contrasts of interest. This is a flexible approach to modeling factorial designs that has many strengths. It is possible to evaluate most any contrast of substantive interest. For example, in Table 16.22, I included but did not discuss a contrast on Lines 88 and 89 that compared the decay in the estimated Y means from the immediate to the follow-up testing periods for the intervention group (`y_tx1_3-y_tx1_2`) to the corresponding mean decay for the control group (`y_tx0_3-y_tx0_2`) using the following `MODEL CONSTRAINT` commands:

```
NEW (decay32) ;
 decay32=(y_tx1_3-y_tx1_2) - (y_tx0_3-y_tx0_2) ;
```

The contrast strategy uses the delta method for estimating standard errors, confidence intervals and significance tests but one can just as easily use bootstrapping or Bayesian approaches as well. A weakness of the strategy is that it cannot evaluate M→Y links but I circumvented this weakness by complementing the analysis with SEM-based fixed effects regression methods. If need be, the latter approach can be extended to more complex scenarios as discussed in the document on web page titled *Additional Applications of SEM-based Fixed Effects Modeling with Panel Data*.

Concluding Comments on Midtreatment Effects

It is not uncommon for trialists to include midtreatment assessments in their RETs. Doing so is advantageous if not for any other reason than it provides an extra time period to incorporate into longitudinal modeling for purposes of exploring M→Y links using sophisticated modeling approaches. Doing so also can permit the analysis of early responders and the strengths of adaptive designs. If an RET study is such that a midtreatment assessment can be included at reasonable cost *and* if one does not believe that including the assessments will introduce testing effects (see Chapter 4), they are well worth considering. There are interesting and sophisticated approaches to identifying responders and non-responders in clinical trials based on a method known as mixture modeling. I discuss these methods in Chapter XX.

CONCLUDING COMMENTS

The current chapter has covered considerable ground on longitudinal analyses in the context of RETs. In many ways, I have only coveredthe tip of the iceberg for each of the focal methods. My goals was to provide you with a working knowledge of key concepts in each area and to provide concrete illustrative examples to start you down the path of analyzing data using a method that most resonates with you. There is no one "correct" way to analyze longitudinal data. Each method comes at the data from different angles and each has strengths and weaknesses, accordingly.

There are forms of longitudinal analysis I did not cover. For example, I did not cover the method of difference-in-differences (DiD) that is becoming increasingly popular in such fields as economics and public health. I did not do so because these methods usually are applied in settings where randomization is not possible and, as such, are beyond the scope of this book. To be sure, they can come into play when selection effects due to dropouts cause randomization to fail. If the types of people dropping out of treatment vary systematically between the treatment and control groups after randomization, DiD is occasionally used to mimic a quasi-experimental correction. However, in the final analysis, they represent a form of quasi-experimental trial design and analysis rather than a formal randomized trial. Sometimes, economists apply DiD to randomized field experiments simply because DiD (via fixed effects regression) is the standard, expected language of their field, even if alternative approaches such as ANCOVA might be statistically more efficient. For a useful introduction to DiD methods, see the on-line course offered by Jeffrey Woolridge.

APPENDIX A: SUPPLEMENTAL PROGRAMMING FOR LGC MODELING

Consider the mediation link TREAT → M → Y, where T is the treatment condition, M is the mediator, and Y is the outcome The link implies two linear equations (using sample notation):

$$Y = a_Y + b_1 M \quad [A.1]$$

$$M = a_M + b_2 TREAT \quad [A.2]$$

Focus first on Equation A.1, namely the equation where Y is the outcome. A well-known property of linear models is that one can substitute into the equation a value of M that is of substantive interest and then knowing the values of a_Y and b_1, one can calculate the (predicted) mean of Y when M equals that value. Suppose I want to know the mean of Y when the mediator equals 0 and the operative linear equation is

$$Y = 1.0 + 2.0 M$$

In this case, the predicted mean of Y when M = 0 is

$$Y = 1.0 + (2.0)(0) = 1.0$$

If I want to know the mean of Y when the mediator equals 3, I calculate

$$Y = 1.0 + (2.0)(3) = 7.0$$

Now suppose I add a second predictor to the equation, as follows:

$$Y = a_Y + b_1 M + b_3 C \quad [A.3]$$

From the above, I can calculate the mean of Y at any combination of values of M and C that are of interest if I know the values a_Y, b_1, and b_3. Suppose the equation is

$$Y = 2.0 + 1.0 M + 3.0 C$$

If I want to know the mean of Y when M = 1 and C = 0, I calculate

$$Y = 2.0 + (1.0)(1.0) + (3.0)(0) = 3.0$$

Suppose C is a covariate. We often want to know what the predicted mean of Y is at different values of M when the covariate is held constant at its mean value, i.e., its "typical" value. Suppose the mean of C is 2.0, then if my interest in the case where M = 1

and C equals its mean, the operative calculations are

$$Y = 2.0 + (1.0)(1.0) + (3.0)(2.0) = 9.0$$

Now, suppose I mean center the covariate. This transformation makes the new mean of $C_{CENTERED} = 0$, which corresponds to 2.0 on the original metric. This transformation changes the intercept of the equation but not the value of b_1. The new equation will be

$$Y = 8.0 + (1.0)(1.0) + (3.0)(0) = 8.0 + (1.0)(1.0) = 9.0$$

Note that when the covariate is mean centered so that its mean is zero, I can essentially ignore it and just calculate the predicted mean from the new equation that multiplies the value of M of interest by b_1 and add to that product the value of a_Y. The result is the covariate adjusted Y mean but I did not have to formally take C into account in my calculations.

It is because of this simplification that researchers sometimes mean center covariates. Doing so has the effect of allowing us to calculate covariate adjusted mean values of Y for different values of M while holding covariates constant at their "typical" values (i.e., their means) but without having to put the non-transformed mean values of C into the equation. I used this strategy when I mean centered the covariates using the `DEFINE` command in Mplus in the LGC syntax for the RET in the main text. In that sense, the equations I then created within the `MODEL CONSTRAINT` command do not have to formally incorporate the covariates. By definition, the covariates have been held constant at their "typical" values.

With this as background, consider Equation A.2, namely $M = a_M + b_2$ TREAT. Using simple algebra it can be shown that

when TREAT = 0, the mean of M will equal a_M (because the term b_2 T now equals zero)

when TREAT =1, the mean of M will equal $a_M + b_2$

I make use of these properties in the `MODEL CONSTRAINT` commands, as I show below.

There is an additional property that I need to mention that concerns the links between the latent intercept factors and the latent Y slope factor to the variables Y1, Y2, Y3 and Y4 in the RET example and the latent M slope factor to M1, M2, M3 and M4. Consider in the RET example the variable M2 which is the mediator at the first six month follow-up. Technically, the model expresses M2 to be a linear function of the latent intercept (MI) and the latent slope (MS) for the mediator as shown in Figure 16.18. This can be expressed in the equation

$$M2 = a_{M2} + L_{M2,MI}\,MI + L_{M2,MS}\,MS$$

where a_{M2} is the measurement intercept for M2, $L_{M2,MI}$ is the loading (or, technically, the path coefficient) that links MI to M2, and $L_{M2,MS}$ is the loading or path coefficient that links MS to M2. When specifying the model in Figure 16.18, rather than estimate the values of a_{M2}, $L_{M2,MI}$ and $L_{M2,MS}$, they were fixed to *a priori* values. By default, Mplus fixes all of the path coefficients from the latent intercept factor for M to the measures M1, M2, M3 and M4 to a value of 1.0, per Figure 16.18. I did not have to do that in the Mplus syntax because Mplus does it internally for me. It does the same for the latent intercept factor for Y to the measures Y1, Y2, Y3, and Y4. Mplus also fixes all of the measurement intercepts to zero in order for the underlying model mathematics to work. Finally, in Lines 10 and 23 of Table 16.13 in the main text, I specified the slope loadings to the right of the pipe in the syntax line to tell Mplus what values to use. The values are 0, 1, 2, and 3 for Y1, Y2, Y3 and Y4 and these same values are used for M1, M2, M3 and M4. I can thus write the equations for M1 through M4 as

$$M1 = 0 + 1*MI + 0*MS = MI \quad [A.4]$$

$$M2 = 0 + 1*MI + 1*MS = MI + 1*MS \quad [A.5]$$

$$M3 = 0 + 1*MI + 2*MS = MI + 2*MS \quad [A.6]$$

$$M4 = 0 + 1*MI + 3*MS = MI + 3*MS \quad [A.7]$$

and for Y1 through Y4 as

$$Y1 = 0 + 1*YI + 0*YS = YI \quad [A.8]$$

$$Y2 = 0 + 1*YI + 1*YS = YI + 1*YS \quad [A.9]$$

$$Y3 = 0 + 1*YI + 2*YS = YI + 2*YS \quad [A.10]$$

$$Y4 = 0 + 1*YI + 3*YS = YI + 3*YS \quad [A.11]$$

I make use of these expressions when I specify equations in the `MODEL CONSTRAINT` commands.

As an example, consider the expression for the predicted value of M3 in the Treatment B condition (which is analogous to the control group in a traditional RET design because it is dummy coded with a value of 0 for the variable TREAT). The equation for M3 from Equation A.6 is

$$M3 = MI + 2*MS$$

I want to substitute in the mean value MI and the mean value MS into the equation but I want those values to be specific to Treatment B. Consider first MI. According to the model in Figure 16.18, the latent intercept factor for the mediator (MI) is a linear function of the treatment condition and the covariate M0, as follows:

$$MI = a_{MI} + p_1\ TREAT + p_2\ M0 \qquad [A.12]$$

Because I mean centered the M0 covariate in the `DEFINE` command, the rightmost term of the above equation drops out, yielding:

$$MI = a_{MI} + p_1\ TREAT \qquad [A.13]$$

For Treatment B, the value of TREAT is equal to zero, so this also eliminates the rightmost term of Equation A.13. The mean MI for Treatment B is simply the intercept a_{MI}:

$$MI_{\text{Mean for Treatment B}} = a_{MI} \qquad [A.14]$$

Next, consider the latent slope factor MS. The model equation for it per Table 16.13 is

$$MS = a_{MS} + p_6\ TREAT \qquad [A.15]$$

Because for Treatment B the value for TREAT is zero, the rightmost term of the above equation drops out and the mean MS is simply a_{MS}.

$$MS_{\text{Mean for Treatment B}} = a_{MS} \qquad [A.16]$$

In my syntax in Table 16.13, I used the label `ma1` for a_{MI} and the label `ma2` for a_{MS}. Referencing back to the core Equation A.6

M3 = MI + 2*MS

and substituting in the mean MI for Treatment B and MS for treatment B, I obtain:

```
M3TB = ma1 + 2*ma2
```

where M3TB is the predicted mean of M at time 3 for Treatment B. This is the command

in the `MODEL CONSTRAINT` section.[21]

Let's do another example using Y4 for Treatment A. The general expression for Y4 from Equation A.11 is

$$Y4 = YI + 3*YS$$

I want to substitute the mean YI and YS for Treatment A into the above expression. Let's first determine the value of YI for Treatment A. The model equation defining the mean of YI based on Table 16.13 is

$$YI = a_{YI} + p_4*MI + p_5*Y0 \quad [A.17]$$

Because the covariate Y0 was mean centered, I can ignore the p_5*Y0 term, yielding

$$YI = a_{YI} + p_4*MI \quad [A.18]$$

To obtain the mean value YI for Treatment A, I need to define the mean MI for Treatment A as an intermediate step in the above Equation. The model equation for MI is

$$MI = a_{MI} + p_1\ TREAT + p_2\ M0 \quad [A.19]$$

which when I eliminate the covariate term given mean centering of covariates becomes

$$MI = a_{MI} + p_1\ TREAT \quad [A.20]$$

For Treatment A, the value of TREAT = 1, so the mean MI for Treatment A is

$$MI_{\text{Mean for Treatment A}} = a_{ML} + p_1 \quad [A.21]$$

Substituting the labels I used in the syntax for $a_{ML} + p_1$, I obtain

$$MI_{\text{Mean for Treatment A}} = ma1 + p1 \quad [A.22]$$

and referring back to Equation A.18, I have

$$YI_{\text{Mean for Treatment A}} = a_{YI} + p_4*(ma1+p1) \quad [A.23]$$

---

[21] In the actual Mplus syntax, I set the term m1tb to equal ma1 and used the expression `M3TB = m1tb + 2*ma2` in place of `M3TB = ma1 + 2*ma2` as a space and time saving device for other parts of the program. However, the two expressions are mathematically equivalent.

I can express the above using labels from my syntax for the right hand side of the expression:

$YI_{\text{Mean for Treatment A}}$ = ya1 + p4*(ma1+p1) [A.24]

Next, I need to determine YS for Treatment A as the second intermediate step. The general equation for YS based on Table 16.13 is

$YS = a_{YS} + p_6*MS$ [A.25]

To apply the above equation, I need to define the value of MS for treatment A. It uses the Equation A.15, which I repeat here for convenience:

$MS = a_{MS} + p_3$ TREAT

Because the value of TREAT = 1 for Treatment A, the mean MS for Treatment A is

$MS_{\text{Mean for Treatment A}} = a_{MS} + p_3$

and using the labels from my Mplus syntax, we obtain

$MS_{\text{Mean for Treatment A}}$ = ma2 + p3

Applying this to Equation A.25, I obtain

$YS_{\text{Mean for Treatment A}} = a_{YS} + p_6*(ma2 + p3)$

and using my syntax labels I obtain

$YS_{\text{Mean for Treatment A}}$ = ya2 + p6*(ma2 + p3)

With the two intermediate terms in place, I return to Equation A.11, which I repeat here for convenience

Y4 = YI + 3*YS

I substitute in my syntax labels with some rearranging to yield

Y4TA = (ya1 + (ma1 + p1)*p4) + (ya2 + (ma2 + p3)*p6)*3;

where Y4TA is the predicted mean of Y at time 4 for Treatment A. Note that from previous calculations ma1+p1 = MI for the treatment condition and (ma2 + p3) = YI for the treatment condition. I can then simplify the above equation to

$Y4TA = (ya1 + MS_{\text{Mean for Treatment A}} *p4) + (YS_{\text{Mean for Treatment A}})*3;$

The above process is not complicated but it is tedious and it is easy to make syntax errors. You can skip it but then you are not able to report some of the supplementary statistics I reported in the main text. You can check your calculations in part by comparing the results for `mdiff1` through `mdiff4` and `ydiff1` through `ydiff4` in the `MODEL CONSTRAINT` section with the results obtained from adding to the `MODEL INDIRECT` section of Table 16.13 statements of the form

```
M1 IND treat ;
M2 IND treat ;
M3 IND treat ;
M4 IND treat ;
Y1 IND treat ;
Y2 IND treat ;
Y3 IND treat ;
Y4 IND treat ;
```

The Case of Multiple Mediators

The above exposition focused on the RET example in the main text that only had a single mediator. The principles are readily extended to the case of multiple mediators but you need to keep a few additional matters in mind. For the two mediator case, let Y be the outcome and MA be one mediator and MB be the second mediator. All three variables are parameterized in the model using growth curves.

In traditional regression modeling with multiple mediators outside of an LGC modeling context, the working equations, absent disturbance terms and using sample notation, are

$Y = a_1 + b_1\ M1 + b_2\ M2$

$MA = a_2 + b_3\ TREAT$

$MB = a_3 + b_4\ TREAT$

When TREAT = 0 (representing the control group), the mean MA for the control group is a_2, the mean MB for the control group is a_3, and the mean of Y for the control group is $a_1 + b_1$ times the mean MA + b_2 times the mean MB.

When TREAT = 1 (representing the intervention group), the mean MA for the intervention group is $a_2 + b_3$, the mean MB for the intervention group is $a_3 + b_4$, and the mean Y for the intervention group is $a_1 + b_1$* mean MA + b_2* mean MB, or stated in terms of the MA and MB expressions, it is $a_1 + b_1* (a_2+b_3) + b_2* (a_3 + b_4)$.

Basically, all of the decomposition rules and algebra follow the same logic as before, but the scenario is yet more tedious and subject to error. However, the extensions are straightforward.

APPENDIX B: CALCULATION OF WITHIN-PERSON AND BETWEEN-PERSON CORRELATIONS

To calculate the observed within-person and between-person correlations, I use the Mplus program with `TYPE = BASIC TWOLEVEL`. Before executing the program, I need to check the output from the main analysis in Table 16.19 where in Line 37 I requested that Mplus save the raw data from the analysis in the folder `datawithlags.dat`. I did this knowing that Mplus would not only output raw data for the variables I used in the analysis but also for the lagged variables that Mplus generated internally. I double check the very end of the output and see the following:

```
Save file
  datawithlags.dat

Order and format of variables

  Y              F10.3
  M1             F10.3
  M2             F10.3
  Y&1            F10.3
  M1&1           F10.3
  M2&1           F10.3
  M1B            F10.3
  M2B            F10.3
  COV1           F10.3
  TREAT          F10.3
  TIME           F10.3
  _NEWTIME       F10.3
  _BINT          F10.3
  ID             I4
```

Mplus refers to lagged variables with an `&1` following the variable name, in this case `Y&1, M1&1 and M2&1`. I use the above information in the `TYPE = BASIC TWOLEVEL` program, whose syntax appears as follows:

```
1.  TITLE: Basic Analyses;
2.  DATA: FILE = datawithlags.dat ;
3.  VARIABLE:
4.     NAMES = Y M1 M2 LY LM1 LM2 M1B M2B COV1
5.       TREAT TIME NEWTIME BINT ID ;
6.     USEVARIABLES = Y M1 M2 LY LM1 LM2 M1B M2B COV1 TREAT ;
7.     WITHIN LM1 LM2 LY ;
8.     BETWEEN M1B M2B COV1 TREAT ;
9.     MISSING * ;
10.    CLUSTER = id;
```

```
11.  ANALYSIS:
12.  TYPE = BASIC TWOLEVEL ;
13.  OUTPUT: Samp ;
```

On Line 2, I replicate the order on the NAMES line to match that of the program that generated the data. Note however that I simplified the names of the lagged variables to LY LM1 LM2. On Line 7, I list the variables that only occur in the within-person model, namely the three lagged variables. On Line 8, I list the variables that only occur in the between-person model. All variables not referenced in these two lines occur at both the within-person and between-person levels. In Line 9, the value for the missing data is set to * which is what Mplus uses by default when it generated the data. We traditionally have used -9999 but now we must use *. It automatically is applied to all variables.

Here is the relevant output which I heavily edited to make it presentable:

```
NOTE:  The sample statistics for within and between refer to the
       maximum-likelihood estimated within and between covariance
       matrices, respectively.
```

ESTIMATED SAMPLE STATISTICS FOR WITHIN

Correlations

	Y	M1	M2	LY	LM1	LM2
Y	1.000					
M1	0.325	1.000				
M2	0.315	0.010	1.000			
LY	0.458	0.122	0.137	1.000		
LM1	0.261	0.415	0.037	0.324	1.000	
LM2	0.248	0.019	0.423	0.340	0.066	1.000

ESTIMATED SAMPLE STATISTICS FOR BETWEEN

Correlations

	M1B	M2B	COV1	TREAT	Y	M1	M2
M1B	1.000						
M2B	0.086	1.000					
COV1	0.011	0.060	1.000				
TREAT	0.006	0.000	0.003	1.000			
Y	0.085	0.049	0.213	0.112	1.000		
M1	0.321	0.057	0.360	0.178	0.224	1.000	
M2	0.042	0.202	0.247	0.237	0.283	0.125	1.000

APPENDIX C: MPLUS SYNTAX FOR NON-DETRENDED MIXED MODEL

This appendix presents the Mplus syntax I used to obtain an estimate of the effect of the mediator M on the outcome Y without detrending the data for time effects. Here is the syntax:

```
TITLE: Mixed model analysis with mediation (Non-Detrended);
DATA: FILE = mixedM.dat ;
VARIABLE:
  NAMES = id time m y treat t1 t2 t3;
  USEVARIABLES = m y ; ! t1, t2, t3 are removed from the analysis
  CLUSTER = id;
ANALYSIS:
  TYPE = TWOLEVEL;  ! RANDOM is not needed
  ESTIMATOR = BAYES;
MODEL:
  %WITHIN%
  y ON m ;   !This is the non-detrended within-person effect
  %BETWEEN%
  y on m ;   !This regresses person-mean Y onto the person-mean M
OUTPUT:  STDYX CINTERVAL(HPD) RESIDUAL TECH4 TECH8;
```

The key result is for y on m in the WITHIN portion of the model. For the sake of completeness, I regressed the person-mean Y across time onto the person-mean M across time.