Running head: MISSING NOT AT RANDOM MODELS

Missing Not At Random Models for Latent Growth Curve Analyses

Craig K. Enders

Arizona State University

Citation:

Enders, C.K. (in press). Missing not at random models for latent growth curve analyses. *Psychological Methods*.

Abstract

The last decade has seen a noticeable shift to missing data handling techniques that assume a missing at random (MAR) mechanism, where the propensity for missing data on an outcome is related to other analysis variables. Although MAR is often reasonable, there are situations where this assumption is unlikely to hold, leading to biased parameter estimates. One such example is a longitudinal study of substance use where participants with the highest frequency of use also have the highest likelihood of attrition, even after controlling for other correlates of missingness. There is a rather large body of literature on missing not at random (MNAR) analysis models for longitudinal data, particularly in the field of biostatistics. Because these methods allow for a relationship between the outcome variable and the propensity for missing data, they require a weaker assumption about the missing data mechanism. This manuscript describes two classic MNAR modeling approaches for longitudinal data, the selection model and the pattern mixture model. To date, these models have been slow to migrate to the social sciences, in part, because they required complicated custom computer programs. These models are now quite easy to estimate in popular structural equation modeling programs, particularly Mplus. The purpose of this manuscript is to describe these MNAR modeling frameworks and to illustrate their application on a real data set. Despite their potential advantages, MNAR-based analyses are not without problems and also rely on untestable assumptions. This paper offers practical advice for implementing and choosing among different longitudinal models.

Missing data handling techniques have received considerable attention in the methodological literature during the last 40 years. This literature has largely discredited most of the simple procedures that have enjoyed widespread use for decades, including methods that discard incomplete cases (e.g., listwise deletion, pairwise deletion) and approaches that impute the data with a single set of replacement values (e.g., mean imputation, regression imputation, last observation carried forward). The last decade has seen a noticeable shift to analytic techniques that assume a missing at random (MAR) mechanism, whereby an individual's propensity for missing data on a variable *Y* is potentially related to other variables in the analysis (or in the imputation model), but not to the unobserved values of *Y* itself (Little & Rubin, 2002; Rubin, 1976). Maximum likelihood estimation and multiple imputation are arguably the predominant MAR-based approaches, although inverse probability weighting methods have gained traction in the statistics literature (e.g., Carpenter, Kenward, & Vansteelandt, 2006; Robins & Rotnitzky, 1995; Scharfstein, Rotnitzky, & Robins, 1999). A number of resources are available to readers who are interested in additional details on these methods (e.g., Carpenter et al., 2006; Enders, 2010; Little & Rubin, 2002; Rotnitzky, 2009; Schafer, 1997; Schafer & Graham, 2002).

Although the MAR mechanism is often reasonable, there are situations where this assumption is unlikely to hold. For example, in a longitudinal study of substance use, it is reasonable to expect participants with the highest frequency of use to have the highest likelihood of attrition, even after controlling for other correlates of missingness. Similarly, in a study that examines quality of life changes throughout the course of a clinical trial for a new cancer medication, it is likely that patients with rapidly decreasing quality of life scores are more likely to leave the study because they die or become too ill to participate. The previous scenarios are characterized by a relationship between the outcome variable (i.e., substance use, quality of life) and the propensity for missing data. This so-called missing not at random (MNAR) mechanism is

problematic because MAR-based analyses are likely to produce biased parameter estimates. Unfortunately, there is no empirical test of the MAR mechanism, so it is generally impossible to fully rule out MNAR missingness. This underscores the need for MNAR analysis methods.

There is a rather large body of literature on MNAR analysis models for longitudinal data, particularly in the field of biostatistics (e.g., Albert & Follmann, 2000, 2009; Diggle & Kenward, 1994; Follmann & Wu, 1995; Molenberghs & Kenward, 2007; Little, 1995, 2009; Verbeke, Molenberghs, & Kenward, 2000; Wu & Bailey, 1989; Wu & Carroll, 1988). This literature addresses a wide variety of substantive applications and includes models for categorical outcomes, count data, and continuous variables, to name a few. Although researchers are sometimes quick to discount MAR-based analyses, MNAR models are not without their own problems. In particular, MNAR analyses rely heavily on untestable assumptions (e.g., normally distributed latent variables), and even relatively minor violations of these assumptions can introduce substantial bias. This fact has led some methodologists to caution against the routine use of these models (Demirtas & Schafer, 2003; Schafer, 2003). A common viewpoint is that MNAR models are most appropriate for exploring the sensitivity of one's results to a variety of different assumptions and conditions. Despite their potential problems, MNAR models are important options to consider, particularly when outcome-related attrition seems plausible. At the very least, these procedures can augment the results from an MAR-based analysis.

Although MNAR analysis models have been in the literature for many years, they have been slow to migrate to the social and the behavioral sciences. To date, most substantive applications have appeared in the medical literature (e.g., Hogan, Roy, & Korkontzelou, 2004; Kenward, 1998; Michiels, Molenberghs, Bijnens, Vangeneugden, & Thijs, 2002). The adoption of any novel statistical procedure is partially a function of awareness but is also driven by software

availability. MNAR analyses were traditionally difficult to implement because they required complicated custom programming. These models are now quite easy to estimate in popular structural equation modeling programs, particularly Mplus (Muthén & Muthén, 1998-2010). Consequently, the purpose of this manuscript is to describe two "classic" MNAR modeling families for longitudinal data – selection models and pattern mixture models – and illustrate their use on a real data set. Methodologists continue to develop MNAR analysis methods, most of which extend the models that I describe in this paper (e.g., Beunckens, Molenberghs, Verbeke, & Mallinckrodt, 2008; Dantan, Proust-Lima, Letenneur, & Jacqmin-Gadda, 2008; Lin, McCulloch, & Rosenheck, 2004; Muthén, Asparouhov, Hunter, & Leuchter, in press; Roy, 2003; Roy & Daniels, 2008; Yuan & Little, 2009). By limiting the scope of this manuscript to classic techniques, I hope to provide readers with the necessary background information for accessing these newer approaches. Muthén et al. (in press) provide an excellent overview of these recent innovations.

The organization of the manuscript is as follows. The paper begins with an overview of Rubin's (1976) missing data theory, including a discussion of how selection models and pattern mixture models fit into Rubin's definition of an MNAR mechanism. After a brief review of growth curve models, I then describe classic selection models and pattern mixture models for longitudinal data. Next, I use a series of data analysis examples to illustrate the estimation and interpretation of the models. The manuscript concludes with a discussion of model selection and sensitivity analyses.

Theoretical Background

Some background information on Rubin's (1976) missing data theory is useful for understanding the rationale behind MNAR analysis models. According to Rubin, the propensity for missing data is a random variable that has a distribution. In practical terms, this implies that

each variable potentially yields a pair of scores: an underlying Y value that may or may not be observed, and a corresponding R value that denotes whether Y is observed or is missing (e.g., R = 0 if Y is observed and R = 1 if Y is missing). Under an MNAR mechanism, the data and the probability of missingness have a joint distribution

$$p(Y_i, R_i|\theta, \phi) \quad (1)$$

where p denotes a probability distribution, Y_i is the outcome variable for case i, R_i is the corresponding missing data indicator, θ is a set of parameters that describes the distribution of Y (e.g., growth model parameters), and ϕ contains parameters that describe the propensity for missing data on Y (e.g., a set of logistic regression coefficients that predict R). Collectively, the parameters of the joint distribution dictate the mutual occurrence of different Y values and missing data.

Under an MAR mechanism, Equation 1 simplifies, and it is unnecessary to estimate the parameters that dictate missingness (i.e., ϕ). For this reason, an MAR mechanism is often referred to as ignorable missingness. In contrast, an MNAR mechanism requires an analysis model that includes all parameters of the joint distribution, not just those that are of substantive interest. In practical terms, this means that the statistical analysis must incorporate a submodel that describes the propensity for missing data (e.g., a logistic regression that predicts R). Both the selection model and the pattern mixture model incorporate a model for R into the analysis, but they do so in different ways.

The selection model and the pattern mixture model factor the joint distribution of *Y* and *R* into the product of two separate distributions. In the selection modeling framework, the joint distribution is

$$p(Y_i, R_i|\theta, \phi) = p(Y_i|\theta)p(R_i|Y_i, \phi) \quad (2)$$

where $p(Y_i|\theta)$ is the marginal distribution of *Y*, and $p(R_i|Y_i, \phi)$ is the conditional distribution of missing data, given *Y*. The preceding factorization implies a two-part model where the marginal distribution corresponds to the substantive analysis (e.g., a growth model), and the conditional distribution corresponds to a regression model that uses *Y* to predict the probability of missing data. The regression of *R* on *Y* is inherently inestimable because *Y* is always missing whenever *R* equals one. The selection model achieves identification by imposing strict distributional assumptions, typically multivariate normality. The model tends to be highly sensitive to this assumption, and even slight departures from normality can produce substantial bias.

In the pattern mixture modeling framework, the factorization reverses the role of *Y* and *R* as follows

$$p(Y_i, R_i|\theta, \phi) = p(Y_i|R_i, \theta)p(R_i|\phi) \quad (3)$$

where $p(Y_i|R_i, \theta)$ is the conditional distribution of *Y*, given a particular value of *R*, and $p(R_i|\phi)$ is the marginal distribution of *R*. The preceding factorization implies a two-part model where the conditional distribution of *Y* represents the substantive model parameters for a group of cases

that shares the same missing data pattern, and the marginal distribution of *R* describes the incidence of different missing data patterns. This factorization implies the following strategy: stratify the sample into subgroups that share a common missing data pattern and estimate the substantive model separately within each pattern. Although it is not immediately obvious, the pattern mixture model is also inestimable without invoking additional assumptions. For example, a growth model is underidentified in a group of cases with only two observed data points, so these assumptions would take the form of assumed values for the inestimable parameters. I discuss these assumptions in detail later in the manuscript, but suffice to say that the model is prone to bias when its assumptions are incorrect.

The selection model and pattern mixture model are equivalent in the sense that they describe the same joint distribution. However, because the two frameworks require different assumptions, they can (and often do) produce very different estimates of the substantive model parameters. There is usually no way to judge the relative accuracy of the two models because both rely heavily on untestable assumptions. For this reason, methodologists generally recommend sensitivity analyses that apply different models (and thus different assumptions) to the same data. I illustrate the application of these models to longitudinal data later in the manuscript.

Brief Overview of Growth Curve Models

Much of the methodological work on MNAR models has centered on longitudinal data analyses, particularly growth curve models (also known as mixed effects models, random coefficient models, and multilevel models). Because this manuscript focuses solely on longitudinal data analyses, a brief overview of the growth curve model is warranted before proceeding. A growth model expresses the outcome variable as a function of a temporal predictor

variable that captures the passage of time. For example, the unconditional linear growth curve model is

$$Y_{ti} = \beta_0 + \beta_1(TIME_{ti}) + b_{0_i} + b_{1_i}(TIME_{ti}) + \varepsilon_{ti} \quad (4)$$

where Y_{ti} is the outcome score for case *i* at time *t*, $TIME_{ti}$ is the value of the temporal predictor for case *i* at time *t* (e.g., the elapsed time since the onset of the study), β_0 is the mean intercept, β_1 is the mean growth rate, b_{0_i} and b_{1_i} are residuals (i.e., random effects) that allow the intercepts and the change rates, respectively, to vary across individuals, and ε_{ti} is a time-specific residual that captures the difference between an individual's fitted linear trajectory and their observed data. The model can readily incorporate non-linear change via polynomial terms. For example, the unconditional quadratic growth model is

$$Y_{ti} = \beta_0 + \beta_1(TIME_{ti}) + \beta_2\left(TIME_{ti}^2\right) + b_{0_i} + b_{1_i}(TIME_{ti}) + b_{2_i}\left(TIME_{ti}^2\right) + \varepsilon_{ti} \quad (5)$$

where β_0 is the mean intercept, β_1 is the average instantaneous linear change when *TIME* equals zero, and β_2 is the mean curvature. As before, the model uses a set of random effects to incorporate individual heterogeneity in the developmental trajectories (i.e., b_{0_i}, b_1 , and b_{2_i}), and ε_{ti} is a time-specific residual.

The previous models are estimable from the multilevel, mixed model, or the structural equation modeling frameworks. Structural equation modeling – and the Mplus software package,

in particular – provides a convenient platform for estimating MNAR models. Cast as a structural equation model, the individual growth components (i.e., b_{0_i}, b_{1_i}, and b_{2_i}) are latent variables, the means of which (i.e., β_0, β_1, and β_2) define the average growth trajectory. To illustrate, Figure 1 shows a path diagram of a linear growth model from a longitudinal study with four equally spaced assessments. The unit factor loadings for the intercept latent variable reflect the fact that the intercept is a constant component of each individual's idealized growth trajectory, and the loadings for the linear latent variable capture the timing of the assessments (i.e., the *TIME* scores in Equation 4). A quadratic growth model incorporates an additional latent factor with loadings equal to the square of the linear factor loadings. A number of resources are available to readers who want additional details on growth curve models (Bollen & Curran, 2006; Hancock & Lawrence, 2006; Hedeker & Gibbons, 2006; Singer & Willett, 2003). As an aside, mixed modeling software programs (e.g., PROC MIXED in SAS) can also estimate some of the MNAR models that I describe in this manuscript (e.g., the selection models). Although different modeling frameworks often yield identical parameter estimates, the latent growth curve approach is arguably more convenient for implementing MNAR models.

Selection Models for Longitudinal Data

Heckman (1976, 1979) originally proposed the selection model as a bias correction method for regression analyses with MNAR data on the outcome variable. Like their classic predecessor, selection models for longitudinal data combine a substantive model (i.e., a growth curve model) with a set of regression equations that predict missingness. The two parts of the model correspond to the factorization on the right side of Equation 2. The literature describes two classes of longitudinal models that posit different linkages between the repeated measures variables and the missing data indicators. The Wu and Carroll (1988) model indirectly links the

repeated measures variables to the response probabilities via the individual intercepts and slopes (i.e., the b_{0_i} and b_{1_i} terms in Equation 4). This approach is commonly referred to as the random coefficient selection model or the shared parameter model[1]. In contrast, the Diggle and Kenward (1994) selection model directly relates the probability of missing data at time *t* to the outcome variable at time *t*. Although these models have commonalities, they require somewhat different assumptions and may produce different estimates. This section provides a brief description of the two models, and a number of resources are available to readers who are interested in additional technical details (Albert & Follmann, 2009; Diggle & Kenward, 1994; Little, 2009; Molenberghs & Kenward, 2007; Verbeke, Molenberghs, & Kenward, 2000).

The Wu and Carroll Model

The Wu and Carroll (WC) model uses the individual growth trajectories to predict the probability of missing data at time *t*. To illustrate, Figure 2 shows a path diagram of a linear WC-type growth curve model. The rectangles labeled R_2, R_3, and R_4 are missing data indicators that denote whether the outcome variable is observed at a particular assessment (e.g., $R_t = 0$ if Y_t is observed and $R_t = 1$ if Y_t is missing). Note that the model does not require an R_1 indicator when the baseline assessment is complete, as is the case in the figure. The dashed arrows that link the latent variables (i.e., the individual intercepts and slopes) to the missing data indicators represent logistic regression equations[2]. Regressing the indicator variables on the intercepts and slopes effectively allows the probability of missing data to depend on the entire set of repeated measures variables, including the unobserved scores from later assessments. Although this proposition may

[1] Authors often treat the shared parameter model as a distinct MNAR approach. Because the structural features of the Wu and Carroll model are similar to those of the Diggle and Kenward model (i.e., one or more variables from the substantive model predict missingness), I treat both as selection models.
[2] A logistic model is not the only possibility for the missing data indicators. Probit models are also common.

seem awkward, linking the response probabilities to the intercepts and slopes is useful when missingness is potentially dependent on an individual's overall developmental trajectory rather than a single error-prone realization of the outcome variable (Albert & Follmann, 2009; Little, 1995).

The Diggle and Kenward Model

The Diggle and Kenward (DK) model also combines a growth curve model with a set of regression equations that predict missingness. However, unlike the WC model, the probability of missing data at wave *t* depends directly on the repeated measures variables. To illustrate, Figure 3 shows a path diagram of a linear DK growth curve model. As before, the rectangles labeled R_2, R_3, and R_4 are missing data indicators that denote whether the outcome variable is observed or missing, and the dashed arrows represent logistic regression equations. Notice that the probability of missing data at time *t* now depends directly on the outcome variable at time *t* as well as on the outcome variable from the preceding assessment (e.g., Y_1 and Y_2 predict R_2, Y_2 and Y_3 predict R_3, and so on).

As an aside, the logistic regression equations in the previous models potentially carry information about the missing data mechanism. For example, in the DK model, a significant path between R_t and Y_t implies an MNAR mechanism because dropout at wave *t* is concurrently related to the outcome. Similarly, a significant association between R_t and Y_{t-1} provides evidence for an MAR mechanism because dropout at time *t* is related to the outcome at the previous assessment. Finally, the absence of any relationship between the outcomes and the missing data indicators is consistent with an MCAR mechanism because dropout is unrelated to the variables in the model. Although it is tempting to use the logistic regressions to make inferences about the missing data mechanism, it is important to reiterate that these associations are only estimable because of strict

distributional assumptions. Consequently, using the logistic regressions to evaluate the missing data mechanism is tenuous, at best.

Selection Model Assumptions

Although it is immediately obvious, longitudinal selection models rely on distributional assumptions to achieve identification, and these distributional assumptions dictate the accuracy of the resulting parameter estimates. For the WC model, identification is driven by distributional assumptions for the random effects (i.e., the individual intercepts and slopes), whereas the DK model requires distributional assumptions for the repeated measures variables. Without these assumptions, the models are inestimable (e.g., in the DK model, the regression of R_t on Y_t is inestimable because Y is always missing whenever R equals one). With continuous outcomes, the typical practice is to assume a multivariate normal distribution for the individual intercepts and slopes or for the repeated measures variables. The WC model additionally assumes that the repeated measures variables and the missing data indicators are conditionally independent, given the random effects (i.e., after controlling for the individual growth trajectories, there is no residual correlation between Y_t and R_t). Collectively, these requirements are difficult to assess with missing data, so the accuracy of the resulting parameter estimates ultimately relies on one or more untestable assumptions.

Coding the Missing Data Indicators

Thus far, I have been purposefully vague about the missing data indicators because the appropriate coding scheme depends on the exact configuration of missing values. The WC and DK models were originally developed for studies with permanent attrition (i.e., a monotone missing data pattern). In this scenario, it makes sense to utilize discrete-time survival indicators, such that R_t takes on a value of zero prior to dropout, a value of one at the assessment where

dropout occurs, and a missing value code at all subsequent assessments (e.g., Muthén & Masyn, 2005; Singer & Willett, 2003). In contrast, when a study has only intermittent missing values, it is reasonable to represent the indicators as a series of independent Bernoulli trials, such that R_t takes on a value of zero at any assessment where Y_t is observed and takes on a value of one at any assessment where Y_t is missing.

Most longitudinal studies have a mixture of sporadic missingness and permanent attrition. One option for dealing with this configuration of missingness is to use discrete-time survival indicators to represent the dropout patterns and code intermittent missing values as though they were observed (i.e., for intermittently missing values, R_t takes on a value of zero). Because intermittent missingness is not treated as a target event, this coding strategy effectively assumes that these values are consistent with an MAR mechanism. A second option for dealing with intermittent missingness and permanent attrition is to create indicators that are consistent with a multinomial logistic regression (Albert & Follmann, 2009; Albert, Follmann, Wang, & Suh, 2002), such that the two types of missingness have distinct numeric codes. I illustrate these various coding strategies in the subsequent data analysis examples.

Pattern Mixture Models for Longitudinal Data

Like the selection model, the pattern mixture approach integrates a model for the missing data into the analysis, but it does so in a very different way. Specifically, a pattern mixture analysis stratifies the sample into subgroups that share the same missing data pattern and estimates a growth model separately within each pattern. For example, in a four-wave study with a monotone missing data pattern, the complete cases would form one pattern, the cases that drop out following the baseline assessment would constitute a second pattern, the cases that leave the study after the second wave form a third pattern, and the cases with missing data at the final

assessment only comprise the fourth pattern. Assuming a sufficient sample size within each pattern, the four missing data groups would yield unique estimates of the growth model parameters. Returning to Equation 3, these pattern-specific estimates correspond to the conditional distribution $p(Y_i|R_i, \theta)$, and the group proportions correspond to $p(R_i|\phi)$.

Although the pattern-specific estimates are often informative, the usual substantive goal is to estimate the population growth trajectory. Computing the weighted average of the pattern-specific estimates yields a marginal estimate that averages over the distribution of missingness. For example, the average intercept from the hypothetical four-wave study is

$$\hat{\bar{\beta}}_0 = \hat{\pi}^{(1)}\hat{\beta}_0^{(1)} + \hat{\pi}^{(2)}\hat{\beta}_0^{(2)} + \hat{\pi}^{(3)}\hat{\beta}_0^{(3)} + \hat{\pi}^{(4)}\hat{\beta}_0^{(4)} \quad (6)$$

where the numeric superscript denotes the missing data pattern, $\hat{\pi}^{(p)}$ is the proportion of cases in missing data pattern p, and $\hat{\beta}_0^{(p)}$ is the pattern-specific intercept estimate. Importantly, a pattern mixture analysis does not automatically produce standard errors for the average estimates because these quantities are a function of the model parameters. Consequently, it is necessary to use the multivariate delta method to derive an approximate standard error (Hedeker & Gibbons, 1997; Hogan & Laird, 1997). Fortunately, performing these additional computations is unnecessary because Mplus can readily compute the average estimates and their standard errors.

As an aside, stratifying cases by missing data pattern is also an old MAR-based strategy that predates current maximum likelihood missing data handling techniques (Muthén, Kaplan, & Hollis, 1987). This so-called multiple group approach used between-pattern equality constraints on the model parameters to trick existing structural equation modeling programs into producing

a single set of MAR-based estimates. Although this procedure closely resembles a pattern mixture model, forcing the missing data patterns to have the same parameter estimates effectively ignores the pattern-specific conditioning that is central to the MNAR factorization in Equation 3.

Model Identification

Although its resemblance to a multiple group analysis makes the pattern mixture model conceptually straightforward, implementing the procedure is made difficult by the fact that one or more of the pattern-specific parameters are usually inestimable. To illustrate, consider a four-wave study that employs a quadratic growth model. The model is identified only for the subgroup of participants with complete data. For cases with two complete observations, the linear trend is estimable but the quadratic coefficient and certain variance components are not. The identification issue is most evident in the subgroup that drops out following the baseline assessment, where neither the linear nor the quadratic coefficients are estimable.

Estimating a pattern mixture model requires the user to specify values for the inestimable parameters, either explicitly or implicitly. Using code variables as predictors in a growth model is one way to accomplish this (Hedeker & Gibbons, 1997, 2006). For example, Hedeker and Gibbons (1997) classified participants from a psychiatric drug trial as completers (cases with data at every wave) or dropouts (cases that left the study at some point after the baseline assessment), and they subsequently included the binary missing data indicator as a predictor of the intercepts and slopes in a linear growth model. A linear model with the missing data indicator as the only predictor would be

$$Y_{ti} = \beta_0 + \beta_1(TIME_{ti}) + \beta_2(DROPOUT_i) + \beta_3(DROPOUT_i)(TIME_{ti}) + b_{0_i} + b_{1_i}(TIME_{ti}) + \varepsilon_{ti} \quad (7)$$

where *DROPOUT* denotes the missing data pattern (0 = completers, 1 = dropouts), β_0 and β_1 are the mean intercept and slope, respectively, for the complete cases, β_2 is the intercept difference for the dropouts, and β_3 is the slope difference for dropouts. The Hedeker and Gibbons (1997, 2006) approach achieves identification by sharing information across patterns. For example, the model in Equation 7 implicitly assumes that early dropouts have the same developmental trajectory as the cases that drop out later in the study. The model also assumes that all missing data patterns share the same covariance structure.

A second estimation strategy is to implement so-called identifying restrictions that explicitly equate the inestimable parameters from one pattern to the estimable parameters from one or more of the other patterns. Later in the manuscript, I illustrate three such restrictions: the complete case missing variable restriction, the neighboring case missing variable restriction, and the available case missing variable restriction. As its name implies, the complete case missing variable restriction equates the inestimable parameters to the estimates from the complete cases. The neighboring case missing variable restriction replaces inestimable parameters with estimates from a group of cases that share a comparable missing data pattern. For example, in a four-wave study, the cases that drop out after the third wave can serve as a donor pattern for the cases that drop out after the second wave, such that the two patterns share the same quadratic coefficient. Finally, the available case missing variable restriction replaces inestimable growth parameters with the weighted average of the estimates from other patterns. Still considering a group of cases with two observations, this identifying restriction would replace the inestimable quadratic term

with the average coefficient from the complete cases and the cases that drop out following the third wave. Additional details and examples of various identification strategies are available elsewhere in the literature (Demirtas & Schafer, 2003; Enders, 2010; Molenberghs, Michiels, Kenward, & Diggle, 1998; Thijs, Molenberghs, Michiels, & Curran, 2002; Verbeke & Molenberghs, 2000).

Pattern Mixture Model Assumptions

The assumed values for the inestimable parameters dictate the accuracy of the pattern mixture model. To the extent that the values are correct, the model can reduce or eliminate the bias from an MNAR mechanism. However, like the selection model, there is ultimately no way to gauge the accuracy of the resulting estimates, and implementing different identification constraints can (and often does) produce disparate sets of results. At first glance, the need to specify values for inestimable parameters may appear to be a serious weakness of the pattern mixture model. However, some methodologists argue that this requirement is advantageous because it forces researchers to make their assumptions explicit. This is in contrast to the selection model, which relies on implicit distributional assumptions that are unobvious. This aspect of the pattern mixture model also provides flexibility because it allows researchers to explore the sensitivity of the substantive model parameters to a number of different identification constraints (i.e., assumed parameter values). In truth, the previous identifying restrictions are simply arbitrary rules of thumb for generating parameter values. Any number of other restrictions is possible (e.g., a restriction that specifies a flat trajectory shape after the last observed data point; Little, 2009), and performing a sensitivity analysis that applies a variety of identification strategies to the same data is usually a good idea.

Data Analysis Examples

To date, applications of longitudinal models for MNAR data are relatively rare in the social and the behavioral sciences, perhaps because these analyses have traditionally required complex custom programming. Software availability is no longer a limiting factor because the Mplus package provides a straightforward platform for estimating a variety of selection models and pattern mixture models. This section describes a series of data analyses that apply the MNAR models from earlier in the manuscript. The Mplus 6 syntax files for the analyses are available at www.appliedmissingdata.com/papers.

The analysis examples use the psychiatric trial data from Hedeker and Gibbons (1997, 2006)[3]. Briefly, the data were collected as part of the National Institute of Mental Health Schizophrenia Collaborative Study and consist of repeated measurements from 437 individuals. In the original study, participants were assigned to one of four experimental conditions (a placebo condition and three drug regimens), but the subsequent analyses collapse these categories into a dichotomous treatment indicator (0 = placebo, 1 = drug). The primary substantive goal is to assess treatment-related changes in illness severity over time. The outcome was measured on a 7-point scale, such that higher scores reflect greater severity (e.g., 1 = normal, not at all ill; 7 = among the most extremely ill). Most of the measurements were collected at baseline, week one, week three, and week six, but a small number of participants also had measurements at week two, week four, or week five. To simplify the presentation, I excluded these irregular observations from all analyses. Finally, note that the discrete measurement scale violates multivariate normality, by definition. Although these data are still useful for illustration purposes, the normality violation is likely problematic for the selection model analyses.

[3] The data are used here with Dr. Hedeker's permission and are available at his website: http://tigger.uic.edu/~hedeker/long.html

The data set contains nine distinct missing data patterns that represent a mixture of permanent attrition and intermittent missingness. The leftmost panel of Table 1 summarizes these patterns. To provide some sense about the developmental trends, Figure 4 shows the observed means for each pattern by treatment condition. The fitted trajectories in the figure suggest non-linear growth. In their analyses of the same data, Hedeker and Gibbons (1997, 2006) linearize the trajectories by modeling illness severity as a function of the square root of weeks. Although this decision is very sensible, I used a quadratic growth model for the subsequent analyses because it provides an opportunity to illustrate the complexities that arise with MNAR models, particularly pattern mixture models with identifying restrictions. The analysis model is

$$\begin{aligned} Y_{ti} = \beta_0 + \beta_1(TIME_{ti}) + \beta_2\left(TIME_{ti}^2\right) + \beta_4(DRUG_i) + \beta_5(DRUG_i)(TIME_{ti}) \\ + \beta_6(DRUG_i)\left(TIME_{ti}^2\right) + b_{0_i} + b_{1_i}(TIME_{ti}) + b_{2_i}\left(TIME_{ti}^2\right) + \varepsilon_{ti} \end{aligned} \tag{8}$$

where β_0, β_1, and β_2 define the average growth trajectory for the placebo cases (i.e., *DRUG* = 0), and β_4, β_5, and β_6 capture the mean differences between the treatment conditions.

In an intervention study, the usual goal is to assess treatment group differences at the end of the study. Centering the temporal predictor at the final assessment (e.g., by fixing the final slope factor loading to a value of 0) addresses this question because the regression of the intercept on treatment group membership quantifies the mean difference. However, implementing identifying restrictions in a pattern mixture model is made easier by centering the intercept at the baseline assessment, particularly when permanent attrition is the primary source of missingness. Consequently, I fixed the linear slope factor loadings to values of 0, 1, 3, and 6 for all subsequent analyses (the quadratic factor loadings are the squares of these values). Despite this

parameterization, it is straightforward to construct a test of the endpoint mean difference. Algebraically manipulating the growth model parameters gives the model-implied mean difference at the final assessment

$$\hat{\mu}_{Drug} - \hat{\mu}_{Placebo} = \hat{\beta}_3 + 6 \cdot \hat{\beta}_4 + 36 \cdot \hat{\beta}_5 \quad (9)$$

where the $\hat{\beta}$ terms are the regression coefficients that link the growth factors to the treatment indicator (i.e., the latent mean differences), 6 is the value of the linear factor loading at the final assessment (i.e., the time score, weeks since baseline), and 36 is the corresponding quadratic factor loading. Among other things, the MODEL CONSTRAINT command in Mplus allows users to define new parameters that are functions of the estimated parameters. In the subsequent analyses, I used this command to estimate the mean difference in Equation 9 and its standard error.

MAR-Based Growth Curve Model

As a starting point, I used MAR-based maximum likelihood missing data handling to estimate the quadratic growth curve model. Figure 5 shows the path diagram for the analysis. Table 2 lists the estimates and the standard errors for selected parameters, and Figure 6 displays the corresponding model-implied trajectories. The figure clearly suggests that participants in the drug condition experienced greater reductions in illness severity relative to the placebo group. However, it is important to emphasize that these estimates assume that an individual's propensity for missing data at week *t* is completely determined by treatment group membership or by his or her severity score at earlier assessments (i.e., the missing values conform to an MAR mechanism).

Substituting the appropriate quantities from the maximum likelihood analysis into Equation 9 gives a mean difference of -1.424 (SE = .182, p < .001). Expressed relative to the model-implied estimate of the baseline standard deviation (i.e., the square root of the sum of the intercept variance and the residual variance), the standardized mean difference is d = 1.563.

Diggle and Kenward Selection Models

Mplus is ideally suited for estimating selection models because it can accommodate normally distributed (e.g., the repeated measures) and categorical outcomes (e.g., the missing data indicators) in the same model. To illustrate, I fit two DK-type selection models to the psychiatric trial data. The first analysis treated permanent attrition (patterns 2 through 4) as MNAR and treated intermittent missingness (patterns 5 through 9) as MAR. As noted previously, missing data indicators that are consistent with a discrete-time survival model are appropriate when modeling permanent dropout. Normally, a set of three missing data indicators could represent the dropout patterns in Table 1, but the small number of cases in pattern 4 made it impossible to model attrition at the second assessment. Consequently, the model incorporated indicator variables at the final two waves with the following coding scheme

$$R_t = \begin{cases} 0 & \text{observed or intermittent missingness} \\ 1 & \text{dropout at time } t \\ 99 & \text{dropout at previous time} \end{cases}$$

where 99 represents a missing value code. Importantly, assigning a code of 0 to the intermittently missing values effectively defines sporadic missingness (patterns 5 though 9) as MAR. Finally, note that the pattern 4 cases had indicator codes of R_3 = 1 and R_4 = 99. This treats the missing Y_2

values are MAR and the missing Y_3 values as MNAR dropout. The middle panel of Table 1 summarizes the indicator codes for each missing data pattern.

Figure 7 shows a path diagram of the DK selection model. The different types of dashed arrows represent equality constraints on the regression coefficients in the logistic part of the model (e.g., the regression of R_4 on Y_4 is set equal to the regression of R_3 on Y_3). Describing the specification of a discrete-time survival model is beyond the scope of this manuscript, but readers that are interested in the rationale behind these constraints can consult Singer and Willett (2003) and Muthén and Masyn (2005), among others.

Table 3 gives selected parameter estimates and standard errors from the analysis. The model-implied growth trajectories were quite similar to those in Figure 6, although the selection model produced a larger mean difference between the treatment conditions at week six. Specifically, substituting the appropriate estimates into Equation 9 yields a model-implied mean difference of -1.665 (SE = .198, p < .001) at the final assessment. Expressed relative to the model-implied estimate of the baseline standard deviation, this mean difference corresponds to a standardized effect size of d = 1.810. Notice that the selection model produced the same substantive conclusion as the maximum likelihood analysis (i.e., the drug condition experienced greater reductions in illness severity), albeit with a larger effect size. Again, the normality violation should cast doubt on the validity of the selection model estimates.

Turning to the logistic portion of the model, the regression coefficients quantify the influence of treatment group membership and the repeated measures variables on the hazard probabilities (i.e., the conditional probability of dropout, given participation at the previous assessment). For example, the significant positive association between R_t and Y_t suggests that participants with higher illness severity scores at wave t were more likely to drop out, even after

controlling for treatment group membership and scores from the previous assessment. Although the accuracy of this coefficient depends on untenable distributional assumptions, it does provide some evidence for an MNAR mechanism. It is important to note that estimating the model from 100 random starting values produced two sets of solutions with different logistic regression coefficients (the log likelihood values were -2565.814 and -2573.115). In the second solution, the association between R_t and Y_t switched signs, such that cases with lower illness severity scores were more likely to drop out. It is unclear whether this sensitivity to different starting values is a symptom of model misspecification (e.g., the logistic portion of the model omits an important predictor of missingness) or normality violation. Because these models are weakly identified to begin with, a quadratic model may be too complex, although a linear model showed similar instability. Regardless of the underlying cause, this finding underscores the importance of using random starting values when estimating these models.

The previous analysis treated intermittent missing values as MAR. As an alternative, creating indicators that are consistent with a multinomial logistic regression can distinguish between intermittent and permanent missing values (Albert & Follmann, 2009; Albert, Follmann, Wang, & Suh, 2002). The coding scheme below is one such example

$$R_t = \begin{cases} 0 & \text{intermittent missingness at time } t \\ 1 & \text{dropout at time } t \\ 2 & \text{observed at time } t \\ 99 & \text{dropout at an earlier time} \end{cases}$$

where 99 is a missing value code. By default, Mplus treats the highest non-missing category (e.g., 2) in a multinomial logistic regression as the reference group, so assigning the highest code to the

observed values yields logistic regression coefficients that quantify the probability of each type of missingness relative to complete data. After minor alterations to accommodate sparse missing data patterns, I estimated the DK model under this alternate coding scheme. The rightmost panel of Table 1 summarizes the indicator coding for the analysis. The model with multinomial indicators produced mean difference and effect size estimates that were quite similar to those of the previous DK model. The logistic portion of the model was also comparable. The similarity of the two coding schemes suggests that treating intermittent missing values as MAR had very little impact on the final estimates, perhaps due to the fact that permanent attrition accounts for the vast majority of the missing data.

Wu and Carroll Selection Model

In the previous DK models, the probability of missing data was directly related to the repeated measures variables. In contrast, the WC selection model uses individual intercepts and slopes as predictors of missingness. Although it is possible to recast both of the previous DK models as WC models, only the model with discrete-time survival indicators converged to a proper solution. Consequently, I limit the subsequent discussion to an analysis that treated permanent attrition (patterns 2 through 4) as MNAR and treated intermittent missingness (patterns 5 through 9) as MAR. The missing data indicators were identical to the discrete-time coding scheme in the middle panel of Table 1 (i.e., 0 = observed or intermittent missingness, 1 = dropout at time *t*, 99 = dropout at a previous time). An initial analysis failed to converge because the latent variable covariance matrix was not positive definite. Constraining the quadratic factor variance to zero eliminated this problem and produced plausible parameter estimates. Because of this modification, the final model used treatment group membership and the individual intercepts and linear slopes to predict attrition. Figure 8 shows a path diagram of the final model.

As before, different types of dashed arrows represent equality constraints on the regression coefficients in the logistic part of the model.

It is important to note that estimating the model from 100 random starting values produced 85 convergence failures, even after eliminating the quadratic variance from the model. The 15 sets of starts that successfully converged produced comparable log likelihood values but slightly different parameter estimates. Simplifying the model by examining change as a linear function of the square root of time reduced this problem and produced sets of solutions with identical estimates and identical log likelihood values. This finding suggests that a quadratic model is too complex for these data, but it could also be the case that model misspecification or normality violations are contributing to the convergence failures. For the illustration purposes, I report the quadratic model estimates from the solution with the highest log likelihood, but these results should be viewed with caution.

Table 4 gives selected parameter estimates and standard errors from the WC selection model. The WC model produced smaller effect size than the DK selection model. Specifically, substituting the appropriate estimates into Equation 9 yields a mean difference of -1.363 (SE = .183, $p < .001$) at the final assessment and a standardized effect size of $d = 1.576$. Turning to the logistic portion of the model, the regression coefficients quantify the influence of the individual intercepts and linear slopes on the hazard probability. Because the time scores are centered at the baseline assessment, the linear slope represents instantaneous change at the beginning of the study. Consequently, the negative coefficient for the regression of R_t on the linear growth factor suggests that participants who experienced immediate reductions in illness severity were most likely to drop out, even after controlling for initial severity level (i.e., the intercept) and treatment group membership.

Overview of the Pattern Mixture Models

Hedeker and Gibbons (1997, 2006) illustrate a pattern mixture modeling approach that uses the missing data pattern (represented by one or more dummy variables) as a predictor in the growth model. This method is advantageous because standard mixed modeling procedures (e.g., the MIXED procedures in SPSS and SAS) can estimate the model. Mplus offers finite mixture modeling options (Muthén & Shedden, 1999) that are ideally suited for implementing a variety of other pattern mixture models that are difficult or impossible to estimate with standard software (e.g., pattern mixture models with identifying restrictions). Because Hedeker and Gibbons thoroughly describe the use of pattern indicators as predictors of growth, I limit the subsequent examples to pattern mixture models with identifying restrictions. Interested readers can consult Muthén et al. (in press) for other interesting variations of the pattern mixture model.

Within the Mplus finite mixture modeling framework, each missing data pattern functions as a pseudo latent class. In the conventional pattern mixture model, these "classes" simply reflect a manifest grouping variable that is derived from the observed missing data patterns. For example, in a simple model, the complete cases could form one class, and the cases with one or more missing values could comprise a second class. The KNOWNCLASS subcommand in Mplus uses a grouping variable from the input data set to assign cases to classes with a probability of zero or one. Although the pattern mixture models in this section are effectively multiple group growth models, the finite mixture modeling framework provides a convenient mechanism for implementing various identifying restrictions (a multiple group model does not allow the user to specify equality constraints for inestimable parameters). Roy (2003) and Muthén et al. (in press) describe modeling variations that treat class membership as a true latent variable.

Returning to the psychiatric trial data, Figure 4 shows the observed means and the fitted trajectories for each of the nine missing data patterns. With a small number of patterns and a sufficiently large sample size, it would be possible to define each pattern as a distinct class, but the number of cases in patterns 4 through 9 precludes this option. To simplify the models, I reduced the number of classes by aggregating patterns with comparable trajectory shapes. Considering the first three patterns, there appears to be a relationship between dropout time and the rate of initial decline, such that rapid improvement is associated with earlier dropout, at least in the drug condition. Consequently, it is reasonable to treat the first three patterns as distinct classes. Although the decision was somewhat arbitrary, I combined patterns 3 and 4 because these groups were comparable with respect to the timing of dropout. Next, consider the cases with intermittent missingness (patterns 5 through 9). Although it is reasonable to treat these patterns as a distinct group, the trajectory shapes roughly resemble the growth curves for the complete cases. Because the BIC values from a series of preliminary analyses clearly favored a model that combined patterns 5 through 9 with pattern 1, the final set of pattern mixture models used three classes: (a) cases with complete data and intermittent missing values, (b) cases that dropped out after the third assessment, and (c) cases that dropped out after the first or the second assessment.

Recall that pattern mixture models are inherently underidentified because they typically involve one or more inestimable parameters. With respect to the mean structure, classes 1 and 2 have sufficient data to estimate a quadratic trend, but the quadratic intercept and the regression of the quadratic growth factor on the treatment group indicator are inestimable for class 3. The subsequent models used one of three identifying restrictions to achieve identification. The complete case missing variable restriction equated the inestimable quadratic parameters to those of class 1 (complete data and intermittent missingness). The second model implemented the

neighboring case missing variable restriction by replacing the inestimable parameters with those of class 2 (dropout after the third assessment). The final model used the available case missing variable restriction and equated the quadratic parameters for class 3 to the weighted average of the estimates from classes 1 and 2. In Mplus, specifying between-class equality constraints (e.g., using the MODEL CONSTRAINT command) implements these restrictions. Although the same identification strategies are applicable to the covariance structure, the subsequent models assumed a common covariance matrix for the three classes.

Complete Case Missing Variable Restriction

Recall that the pattern mixture model produces unique parameter estimates for each class (i.e., estimates that are conditional on the missing data pattern). Although the substantive goal is to generate a single set of estimates that averages across the distribution of missing data, it is important to inspect the class-specific results. To better illustrate the estimates, panel A of Figure 9 shows the model-implied growth curves for each class. Notice that the fitted trajectories for the class 3 drug condition and the class 2 placebo condition fall outside the plausible score range. For class 3, the identifying restriction clearly underestimated the degree of curvature. For class 2, the mean structure was identified, but attrition at the final assessment produced an inaccurate extrapolation. After some experimentation, changing the constrained value of the class 3 regression coefficient from .021 to .080 produced a reasonable trajectory that stayed within bounds. Similarly, constraining the class 2 intercept to a value of .070 or lower returned plausible estimates.

At first glance, it may seem unreasonable to arbitrarily change parameter values. However, it is important to remember that the identifying constraints essentially represent assumptions about trajectory shapes that would have been observed, had the data been complete.

Because the growth curves in panel A clearly represent incorrect predictions about the unobserved data points, it is difficult to defend a set of marginal estimates that average across the missing data patterns. Consequently, model modification seems necessary in this case. In truth, the identifying restrictions are nothing more than arbitrary rules of thumb for generating plausible parameter values, so viewing the restrictions as tentative starting points for estimation and altering them as needed is a sensible strategy.

After implementing new parameter constraints, the model produced plausible class-specific estimates. The top section of Table 5 gives the updated estimates, and panel B of Figure 9 displays the corresponding model-implied trajectories. Computing the weighted mean of the class-specific values yields an estimate of the population growth trajectory that averages over the distribution of missingness. For these analyses, the population estimate is

$$\hat{\bar{\theta}} = \hat{\pi}^{(1)}\hat{\theta}^{(1)} + \hat{\pi}^{(2)}\hat{\theta}^{(2)} + \hat{\pi}^{(3)}\hat{\theta}^{(3)} \tag{10}$$

where the numeric superscript denotes the missing data pattern, $\hat{\pi}^{(p)}$ is the proportion of cases in missing data class p, and $\hat{\theta}^{(p)}$ is the class-specific estimate. Because the averaging process is identical for all estimates, $\hat{\theta}$ generically denotes a model parameter.

Table 6 gives the average estimates and the standard errors for selected parameters. The trajectory shapes from the pattern mixture model resemble those from the previous analyses, but the mean difference at the final assessment is somewhat larger. Specifically, using the class-specific coefficients to construct a mean difference for each missing data pattern and computing the weighted average of these estimates gives a difference of -1.827 (SE = .374, p < .001) at the final

assessment. Expressed relative to the model-implied estimate of the baseline standard deviation, this mean difference equates to a standardized effect size of $d = 2.019$. Because the marginal estimates (i.e., the result of Equation 10) are a function of the model parameters, a pattern mixture analysis does not automatically produce standard errors. Consequently, it is necessary to use the multivariate delta method to derive an approximate standard error (Hedeker & Gibbons, 1997; Hogan & Laird, 1997). Fortunately, the Mplus MODEL CONSTRAINT command can generate the average estimates and their standard errors, so further computations are unnecessary. Descriptions of the multivariate delta method are available elsewhere in the literature (MacKinnon, 2008; Raykov & Marcoulides, 2004), and Enders (2010) sketches the computational details for various identifying restrictions.

Neighboring Case Missing Variable Restriction

The second pattern mixture model analysis used the neighboring case missing variable restriction to equate the inestimable quadratic parameters for class 3 (dropout after the second assessment) to the estimates from class 2 (dropout after the third assessment). Consistent with the complete case restriction, the initial estimates produced fitted trajectories that fell outside the plausible score range. Panel C of Figure 9 shows the model-implied growth trajectories from the initial analysis. Some experimentation revealed that constraining the quadratic intercept for pattern 2 (and by extension, the quadratic intercept for pattern 3) to a value of .070 or lower produced growth curves that stayed within bounds. The middle portion of Table 5 lists the class-specific estimates from the revised model, and panel D of Figure 9 displays the corresponding trajectories.

Table 6 gives the average estimates and the standard errors from the neighboring case missing variable restriction. The model-implied mean difference is -1.957 ($SE = .522$, $p < .001$), and

the corresponding effect size is d = 2.163. The effect size difference is largely due to the elevated growth trajectory for the placebo condition in class 3. Again, it is important to reiterate that the differences between the two models result from applying different sets of assumptions about the unobserved data. There is no way to empirically assess the accuracy of competing estimates.

Available Case Missing Variable Restriction

The final analysis implemented the available case missing variable restriction. Recall that this approach achieves identification by equating an inestimable parameter to the weighted average of the estimates from other patterns. Applied to the current example, the available case restriction replaced the quadratic intercept for class 3 (dropout after the second assessment) with the weighted average of the intercept estimates from the first two classes. The weight for class 1 was 336/389 = .864, and the weight for class 2 was 53/389 = .136. Consistent with the previous analyses, the initial model produced trajectories that fell outside the plausible score range (see Panel E of Figure 9). Because the available case restriction applied the largest weight to the estimates from the complete cases, the growth curves in panel E closely resemble those from the complete case missing variable restriction in panel A. Changing class 1's contribution to the inestimable regression coefficient from .021 to .080 and constraining the quadratic intercept for class 2 to a value of .070 or lower produced plausible growth curves. Notice that these are the same modifications from the previous analysis. The bottom section of Table 5 gives the class-specific estimates from the revised model, and panel F of Figure 9 displays the corresponding trajectories.

Table 6 displays the population estimates and their standard errors. Perhaps not surprisingly, the available case restriction produced estimates that were virtually identical to those of the complete case missing variable restriction. The similarity owes to the fact that

complete cases primarily determined the values of the inestimable parameters (the weight for this group was .864, as compared to .136 for class 2). The mean difference and standardized effect size values from the analysis (-1.845 and 2.038, respectively) were also virtually identical to those of the complete case restriction (-1.827 and 2.019, respectively).

Analysis Summary

The preceding analysis examples applied seven different models – and thus seven sets of assumptions – to the psychiatric trial data. Although the analyses produced the same substantive conclusion (i.e., the drug group exhibited dramatic improvement relative to the placebo group), the standardized effect size estimates had a range of nearly seven-tenths of a standard deviation unit. Because the models applied different assumptions, this variation might not come as a surprise. Nevertheless, the fluctuation in the effect size estimates is disconcerting. Had the intervention effect not been so dramatic, it could have easily been the case that the models produced conflicting evidence about the efficacy of the drug condition. Unfortunately, it is relatively common for sensitivity analyses to produce discrepant estimates (Demirtas & Schafer, 2003; Foster & Fang, 2003). The next section offers some practical advice on model selection.

Choosing Among Competing Models

MNAR modeling is an active area of methodological research, and the procedures in this manuscript represent just a few possible options. Given the wide array of analytic choices, model selection becomes an important practical consideration; this is particularly true when different models produce disparate estimates, as they do in the preceding examples. Although somewhat disconcerting, it is impossible to provide general recommendations about model selection because every analytic option – MAR or MNAR – relies on one or more untestable assumptions. Although an MAR and an MNAR model may produce identical fit to the observed data, they make

fundamentally different predictions about the unobserved score values (Molenberghs & Kenward, 2007). Because there is no way to empirically assess the validity of these predictions, model selection is not about choosing a single "correct" model. Rather, researchers must choose the model with the most defensible set of assumptions and construct a logical argument that defends that choice. In some situations, it is possible to discount certain models a priori (e.g., the preceding selection model analyses are suspect due to the normality violations). In other situations, substantive considerations may lead researchers to prefer one model over the other. This section outlines a few such considerations.

To begin, consider the selection modeling framework. Although the WC and DK models have commonalities, study-specific features may influence model selection. To illustrate, consider two hypothetical research scenarios. First, suppose that a psychologist is studying quality of life in a clinical trial for a new cancer medication and finds that a number of patients become so ill (i.e., their quality of life becomes so poor) that they can no longer participate in the study. In this situation, it is reasonable to believe that attrition is related to one's developmental trajectory, such that patients with rapidly decreasing quality of life scores are most likely to leave the study because they die or become too ill to participate. To the extent that this assumption is correct, the WC model may be preferred because the developmental trajectories – as opposed to single realizations of the quality of life measure – are probable determinants of missingness. Methodologists have also suggested that the random coefficient model is well-suited for situations where the outcome measure is highly variable over time (Albert & Follmann, 2009) or is an unreliable indicator of an underlying latent construct (Little, 1995).

As a second example, consider a drug treatment study that tracks substance use in the weeks following an intervention. In this situation, it seems plausible that attrition is related to

the actual outcome at time *t*, such that participants who use drugs prior to an assessment fail to show up because they will screen positive for substance use. The DK model may be most appropriate for this scenario because the outcome scores at a particular time point – as opposed to the developmental trends – are likely to determine missingness. Although the substantive research problem may favor one selection model over the other, it is important to reiterate that the data provide no basis for empirically comparing the two models. Consequently, conducting a sensitivity analysis that fits both models to the same data is usually a good strategy.

Substantive and practical considerations also come into play with pattern mixture models. The idea of estimating developmental trajectories separately for each missing data pattern is intuitively appealing, particularly for researchers who are familiar with multiple group structural equation models. In some situations, the class-specific estimates can provide additional insight into one's substantive hypotheses. For example, in an intervention study, it may be interesting to examine the response to treatment within each dropout class in addition to estimating a marginal treatment effect that averages over missing data patterns. Although the previous analysis examples did not illustrate this possibility, the pattern mixture model can incorporate predictors of dropout class membership. This too can provide useful substantive information (e.g., by identifying factors that are related to dropout or to a particular developmental trajectory). One of the pattern mixture model's often-cited advantages is that it forces researchers to explicitly state their assumptions in the form of values for the inestimable parameters. The identifying restrictions that I implemented in the earlier analysis examples are just a few possibilities, and experimenting with different options is quite easy in Mplus. The ability to identify the members of each missing data pattern is potentially useful in this regard. For example, if the members of a particular dropout group share a common set of characteristics (e.g., in a school-based study, the early dropout class has a high proportion of learning disabled children), it might be possible to

use previous research or substantive knowledge to formulate reasonable predictions for the inestimable parameters. The flexibility of the pattern mixture model makes it a highly useful tool for conducting sensitivity analyses.

Sensitivity Analyses

In the missing data literature, a common viewpoint is that researchers should explore the stability of their substantive conclusions by fitting alternate models to the same data. I previously illustrated this procedure by fitting seven different models to the psychiatric trial data. Exploring alternate models is just form of sensitivity analysis, and methodologists have outlined many other procedures. Although it is impossible to briefly summarize the broad range of viewpoints and analytic approaches from the sensitivity analysis literature, it is nevertheless important to raise awareness of this topic. Molenberghs and colleagues (Molenberghs & Kenward, 2007; Molenberghs, Verbeke, & Kenward, 2009) provide a detailed discussion of these procedures, and this section summarizes a few of their key points.

Within a given modeling framework, it is useful to explore the sensitivity of key parameter estimates to various model modifications. As an example, consider the selection modeling framework. Both the DK and WC models are sensitive to minor violations of distributional assumptions; the former assumes that the repeated measures variables are multivariate normal, and the latter assumes that the random effects (i.e., the individual intercepts and slopes) are normal. Examining the change in key parameter estimates after modifying distributional assumptions is an important type of sensitivity analysis. Although it is not the only method for doing so, finite mixture modeling (e.g., growth mixtures) is a useful tool for representing nonnormal manifest variables as well as nonnormal random effects (McLachlan & Peel, 2000; Muthén & Asparouhov, 2009). In the context of MNAR analyses, methodologists have outlined

latent class versions of the DK- and WC-type selection models (Beunckens et al., 2008; Muthén et al., in press) that are readily estimable with Mplus. Similar strategies are available for pattern mixture models (Muthén et al., in press; Roy, 2003; Roy & Daniels, 2008).

Modifying the growth model's covariance structure is second option for exploring sensitivity within a given modeling framework. Conventional wisdom suggests that modifying the covariance structure has little to no impact on average growth rate estimates (Singer & Willett, 2003). In large part, this is due to the fact that the mean and the covariance structure are independent in a complete-data maximum likelihood analysis (i.e., the off-diagonal elements in the parameter covariance matrix equal zero). Because this independence is lost with missing data, modifying the covariance structure (e.g., estimating class-specific variance components; estimating residual covariances; introducing an alternate covariance structure) can potentially alter the latent variable means; Molenberghs et al. (2009) give an example that illustrates this point. Although it is unclear whether these modifications materially affect the performance of MNAR models, they are nevertheless easy to implement.

Finally, methodologists have developed local influence statistics that attempt to identify cases that unduly impact the parameters of the missingness model (e.g., the logistic regressions from the DK model) or the substantive model. These statistics are conceptually similar to familiar measures from the ordinary least squares regression literature (e.g., Cooks D). Although these influence statistics do not necessarily identify respondents with an MNAR missingness mechanism, they can provide important insight into the behavior of a model. For example, there is evidence to suggest that a complete case with an anomalous score profile can influence estimates in a way that gives credence to an MNAR mechanism (Jansen, Hens, Molenberghs, Aerts, Verbeke, & Kenward, 2006; Kenward, 1998). Interested readers can consult various work by

Molenberghs and colleagues for a detailed overview of local influence measures for missing data analyses (Jansen et al., 2006; Molenberghs & Kenward, 2007; Molenberghs et al., 2009).

Discussion

Methodologists have long advocated for the use of MAR-based missing data handling procedures. The MAR assumption is often very reasonable, but there are many situations where missingness is related to the outcome variable itself. This so-called MNAR mechanism is problematic because MAR-based procedures will produce biased estimates. MNAR analysis models have received considerable attention in the biostatistics literature, particularly in the context of longitudinal data. Although some of these models have been in the literature for many years, they have been slow to migrate to the social and the behavioral sciences. The purpose of this manuscript is to describe two classic MNAR modeling frameworks, the selection model and pattern mixture model. The commonality among MNAR models is that they integrate a submodel that describes the propensity for missing data into the analysis. The selection model augments the growth curve analysis with a set of logistic regressions that describe the probability of missing data at each occasion. The pattern mixture approach estimates the growth model separately within each missing data pattern and subsequently averages over the missing data patterns.

The fundamental problem with missing data analyses is that it is generally impossible to fully rule out MNAR missingness; by the same token, it is impossible to disprove the MAR assumption. Despite their intuitive appeal, MNAR analyses rely on untestable assumptions (e.g., normally distributed latent variables, accurate values for inestimable parameters), and relatively minor violations of these assumptions can introduce substantial bias. The fact that MNAR models produce accurate estimates under a relatively narrow range of conditions has led some

methodologists to caution against their routine use. A common opinion is that these models are most appropriate for sensitivity analyses that apply different models (and thus different assumptions) to the same data.

MNAR analysis techniques continue to receive a great deal of attention in the methodological literature and they are likely to gain in popularity. Despite their limitations, these models are important options to consider, particularly when outcome-related attrition seems plausible. At the very least, MNAR models can augment the results from an MAR-based analysis. Although sensitivity analyses are useful for exploring the impact of modeling choices on key parameter estimates, the observed data provide no basis for model selection. Ultimately, choosing a missing data handling technique – be it MAR or MNAR – is really a matter of choosing among a set of competing assumptions. Consequently, researchers should choose a model with the most defensible set of assumptions and they should provide a logical argument that supports this choice.

References

Albert, P.S., & Follmann, D.A. (2000). Modeling repeated count data subject to informative dropout. *Biometrics*, *56*, 667-677.

Albert, P.S., & Follmann, D.A. (2009). Shared-parameter models. In G. Fitzmaurice, M. Davidian, G. Verbeke, & G. Molenberghs (Eds.), *Longitudinal data analysis* (pp. 433-452). Boca Raton, FL: Chapman & Hall.

Albert, P.S., Follmann, D.A., Wang, S.A., & Suh, E.B. (2002). A latent autoregressive model for longitudinal binary data subject to informative missingness. *Biometrics*, *58*, 631-642.

Beunckens, C., Molenberghs, G., Verbeke, G. & Mallinckrodt, C. (2008). A latent-class mixture model for incomplete longitudinal Gaussian data. *Biometrics*, *64*, 96-105.

Bollen, K.A., Curran, P.J. (2006). *Latent curve models: A structural equation approach*. Hoboken, NJ: Wiley.

Carpenter, J.R., Kenward, M.G., & Vansteelandt, S. (1996). A comparison of multiple imputation and doubly robust estimation for analyses with missing data. *Journal of the Royal Statistical Society, Series A*, *169*, 571-584.

Dantan, E., Proust-Lima, C., Letenneur, L., & Jacqmin-Gadda, H. (2008). Pattern mixture models and latent class models for the analysis of multivariate longitudinal data with informative dropouts. *The International Journal of Biostatistics*, *4*, 1-26.

Demirtas, H., & Schafer, J.L. (2003). On the performance of random-coefficient pattern-mixture models for non-ignorable drop-out. *Statistics in Medicine*, *22*, 2553-2575.

Diggle, P., & Kenward, M.G. (1994). Informative dropout in longitudinal data analysis. *Applied Statistics*, *43*, 49-94.

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

Follmann, D., & Wu, M. (1995). An approximate generalized model with random effects for informative missing data. *Biometrics*, *51*, 151-168.

Foster, E.M., & Fang, G.Y. (2004). Alternative methods for handling attrition: An illustration using data from the Fast Track evaluation. *Evaluation Review*, *28*, 434-464.

Hancock, G.R., & Lawrence, F.R. (2006). Using latent growth models to evaluate longitudinal change. In G.R. Hancock & R.O. Mueller (Eds.), *Structural equation modeling: A second course* (pp. 171-196). Greenwood, CT: Information Age Publishing.

Heckman, J.T. (1976). The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models. *The Annals of Economic and Social Measurement*, *5*, 475-492.

Heckman, J.T. (1979). Sample selection bias as a specification error. *Econometrica*, *47*, 153-161.

Hedeker, D., & Gibbons, R.D. (1997). Application of random-effects pattern-mixture models for missing data in longitudinal studies. *Psychological Methods*, *2*, 64-78.

Hedeker, D., & Gibbons, R.D. (2006). *Longitudinal data analysis*. Hoboken, NJ: Wiley.

Hogan, J.W., & Laird, N.M. (1997). Mixture models for the joint distribution of repeated measures and event times. *Statistics in Medicine*, *16*, 239-257.

Hogan, J.W., Roy, J., & Korkontzelou, C. (2004). Handling drop-out in longitudinal studies. *Statistics in Medicine*, *23*, 1455-1497.

Jansen, I., Hens, N., Molenberghs, G., Aerts, M., Verbeke, G., & Kenward, M.G. (2006). The nature of sensitivity in missing not at random models. *Computational Statistics and Data Analysis*, *50*, 830-858.

Kenward, M.G. (1998). Selection models for repeated measurements with non-random dropout: An illustration of sensitivity. *Statistics in Medicine*, *17*, 2723-2732.

Lin, H., McCulloch, C.E., & Rosenheck, R.A. (2004). Latent pattern mixture models for informative intermittent missing data in longitudinal studies. *Biometrics*, *60*, 295-305.

Little, R. (2009). Selection and pattern mixture models. In G. Fitzmaurice, M. Davidian, G. Verbeke, & G. Molenberghs (Eds.), *Longitudinal data analysis* (pp. 409-431). Boca Raton, FL: Chapman & Hall.

Little, R.J.A. (1995). Modeling the drop-out mechanism in repeated-measures studies. *Journal of the American Statistical Association*, *90*, 1112-1121.

Little, R.J.A., & Rubin, D.B. (2002). *Statistical analysis with missing data* (2nd Ed.). Hoboken, NJ: Wiley.

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

McLachlan, G., & Peel, D. (2000). *Finite mixture models*. New York: Wiley.

Michiels, B., Molenberghs, G., Bijnens, L., Vangeneugden, T., & Thijs, H. (2002). Selection models and pattern-mixture models to analyse longitudinal quality of life data subject to drop-out. *Statistics in Medicine*, *21*, 1023-1041.

Molenberghs, G., & Kenward, M.G. (2007). *Missing data in clinical studies*. West Sussex, England: Wiley.

Molenberghs, G., Verbeke, G., & Kenward, M.G. (2009). Sensitivity analysis for incomplete data. In G. Fitzmaurice, M. Davidian, G. Verbeke, & G. Molenberghs (Eds.), *Longitudinal data analysis* (pp. 501-551). Boca Raton, FL: Chapman & Hall.

Molenberghs, G., Michiels, B., Kenward, M.G., & Diggle, P.J. (1998). Monotone missing data and pattern-mixture models. *Statistica Neerlandica*, *52*, 153-161.

Muthén, B., Asparouhov, T. (2009). Growth mixture modeling: Analysis with non-Gaussian random effects. In G. Fitzmaurice, M. Davidian, G. Verbeke, & G. Molenberghs (Eds.), *Longitudinal data analysis* (pp. 144-165). Boca Raton, FL: Chapman & Hall.

Muthén, B., Asparouhov, T., Hunter, A., & Leuchter, A. (in press). Growth modeling with non-ignorable dropout: Alternative analyses of the STAR*D antidepressant trial. *Psychological Methods*.

Muthén, B., Kaplan, D., & Hollis, M. (1987). On structural equation modeling with data that are not missing completely at random. *Psychometrika*, *52*, 431-462.

Muthén, B.O., & Masyn, K. (2005). Discrete-time survival mixture analysis. *Journal of Educational and Behavioral Statistics*, *20*, 27-58.

Muthén, L.K., & Muthén, B.O. (1998-2010). Mplus user's guide (6th Ed.). Los Angeles, CA: Muthén & Muthén.

Muthén, B., & Shedden, K. (1999). Finite mixture modeling with mixture outcomes using the EM algorithm. *Biometrics*, *55*, 463-469.

Raykov, T., & Marcoulides, G.A. (2004). Using the delta method for approximate interval estimation of parameter functions in SEM. *Structural Equation Modeling: A Multidisciplinary Journal*, *11*, 621-637.

Robins, J.M., & Rotnitzky, A. (1995). Semiparametric efficiency in multivariate regression models with missing data. *Journal of the American Statistical Association*, *90*, 122-129.

Rotnitzky, A. (2009). Inverse probability weighted methods. In G. Fitzmaurice, M. Davidian, G. Verbeke, & G. Molenberghs (Eds.), *Longitudinal data analysis* (pp. 453-476). Boca Raton, FL: Chapman & Hall.

Roy, J. (2003). Modeling longitudinal data with nonignorable dropout using a latent dropout class model. *Biometrics*, *59*, 829-836.

Roy, J., & Daniels, M.J. (2008). A general class of pattern mixture models for nonignorable dropout with many possible dropout times. *Biometrics*, *64*, 538-545.

Rubin, D. B. (1976). Inference and missing data. *Biometrika*, *63*, 581-592.

Schafer, J.L. (2003). Multiple imputation in multivariate problems when the imputation and analysis models differ. *Statistica Neerlandica*, *57*, 19-35.

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

Schafer, J.L., & Graham, J.W. (2002). Missing data: Our view of the state of the art. *Psychological Methods*, *7*, 147-177.

Scharfstein, D.O., Rotnitzky, A., & Robins, J.M. (1999). Adjusting for nonigorable drop-out using semi-parametric nonresponse models. *Journal of the American Statistical Association*, *94*, 1096-1146.

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

Thijs, H., Molenberghs, G., Michiels, B., & Curran, D. (2002). Strategies to fit pattern-mixture models. *Biostatistics*, *3*, 245-265.

Verbeke, G., & Molenberghs, G., & Kenward, M.G. (2000). *Linear mixed models for longitudinal data*. New York: Springer-Verlag.

Wu, M.C., & Bailey, K.R. (1989). Estimation and comparison of changes in the presence of informative right censoring: Conditional linear model. *Biometrics*, *45*, 939-955.

Wu, M.C., & Carroll, R.J. (1988). Estimation and comparison of changes in the presence of informative right censoring by modeling the censoring process. *Biometrics*, *44*, 175-188.

Yuan, Y., & Little, R.J.A. (2009). Mixed-effect hybrid models for longitudinal data with nonignorable dropout. *Biometrics*, *65*, 478-486.

Table 1

Missing Data Patterns and Indicator Codes for Data Analysis Examples

		Repeated Measures				Dropout Codes		Multinomial Codes		
Pattern	*n*	Y_1	Y_2	Y_3	Y_4	R_3	R_4	R_2	R_3	R_4
1	312	O	O	O	O	0	0	2	2	2
2	53	O	O	O	M	0	1	2	2	1
3	45	O	O	M	M	1	99	2	1	1
4	3	O	M	M	M	1	99	2	1	1
5	1	O	M	O	M	0	1	0	2	1
6	13	O	O	M	O	0	0	2	0	2
7	2	O	M	M	O	0	0	0	0	2
8	5	O	M	O	O	0	0	0	2	2
9	3	M	O	O	O	0	0	2	2	2

Note. O = observed, M = missing. For dropout codes, 0 = observed, 1 = dropout, and 99 = a missing value code. For multinomial coding, 0 = intermittent missingness, 1 = dropout, and 2 = observed.

Table 2

MAR-Based Maximum Likelihood Estimates

Parameter	*Est.*	*SE*	*p*
Placebo Intercept	5.293	0.083	< .001
Placebo Linear	-0.226	0.083	.001
Placebo Quadratic	0.013	0.013	.210
Intercept Difference	-0.023	0.096	.811
Linear Difference	-0.481	0.095	< .001
Quadratic Difference	0.041	0.014	.001
Week 6 Difference	-1.424	0.181	< .001

Table 3

DK Selection Model Estimates with MNAR Dropout

Parameter	*Est.*	*SE*	*p*
Placebo Intercept	5.259	0.081	< .001
Placebo Linear	-0.137	0.071	.052
Placebo Quadratic	0.014	0.011	.199
Intercept Difference	-0.011	0.094	.905
Linear Difference	-0.509	0.087	< .001
Quadratic Difference	0.039	0.014	.004
Week 6 Difference	-1.665	0.198	< .001
$Y_t \rightarrow R_t$	2.266	0.531	< .001
$Y_{t-1} \rightarrow R_t$	-1.749	0.388	< .001
Treatment $\rightarrow R_t$	0.347	0.358	.333

Table 4

WC Selection Model Estimates with MNAR Dropout

Parameter	*Est.*	*SE*	*p*
Placebo Intercept	5.274	0.080	< .001
Placebo Linear	-0.199	0.051	< .001
Placebo Quadratic	0.002	0.009	.791
Intercept Difference	-0.05	0.094	.599
Linear Difference	-0.435	0.068	< .001
Quadratic Difference	0.036	0.011	.001
Week 6 Difference	-1.363	0.183	< .001
Intercepts → R_t	0.482	0.437	.271
Linear Slopes → R_t	-5.825	2.681	.030
Treatment → R_t	-3.458	1.622	.033

Table 5

Class-Specific Estimates from Pattern Mixture Models

Parameter	Class 1 (n = 336)	Class 2 (n = 53)	Class 3 (n = 48)
Complete Case Identifying Restriction			
Placebo Intercept	5.154	5.456	5.722
Placebo Linear	-0.278	-0.276	-0.216
Placebo Quadratic	*0.021*	**0.070**	**0.021**
Intercept Difference	0.125	-0.250	-0.356
Linear Difference	-0.342	-0.878	-1.076
Quadratic Difference	0.021	0.040	**0.080**
Neighboring Case Identifying Restriction			
Placebo Intercept	5.154	5.456	5.722
Placebo Linear	-0.278	-0.276	-0.264
Placebo Quadratic	0.021	***0.070***	**0.070**
Intercept Difference	0.125	-0.250	-0.356
Linear Difference	-0.342	-0.878	-1.037
Quadratic Difference	0.021	*0.040*	**0.040**
Available Case Identifying Restriction			
Placebo Intercept	5.154	5.456	5.722
Placebo Linear	-0.278	-0.276	-0.222
Placebo Quadratic	*0.021*	***0.070***	**0.028**
Intercept Difference	0.125	-0.250	-0.356
Linear Difference	-0.342	-0.878	-1.071
Quadratic Difference	0.021	*0.040*	**0.075**

Note. Italic typeface denotes donor estimates for class 3, bold typeface denotes constrained parameters.

Table 6

Pattern Mixture Model Estimates Averaged Across Missing Data Patterns

	CCMV		NCMV		ACMV	
Parameter	*Est.*	*SE*	*Est.*	*SE*	*Est.*	*SE*
Placebo Intercept	5.253	0.079	5.253	0.079	5.253	0.079
Placebo Linear	-0.271	0.069	-0.276	0.068	-0.271	0.069
Placebo Quadratic	0.027	0.010	0.033	0.008	0.028	0.009
Intercept Difference	0.027	0.092	0.027	0.092	0.026	0.092
Linear Difference	-0.488	0.093	-0.484	0.096	-0.487	0.093
Quadratic Difference	0.030	0.014	0.026	0.020	0.029	0.014
Week 6 Difference	-1.827	0.374	-1.957	0.522	-1.845	0.388

Note: CCMV = complete case missing variable restriction, NCMV = neighboring case missing variable restriction, ACMV = available case missing variable restriction.

Figure 1. Path diagram of linear growth model.

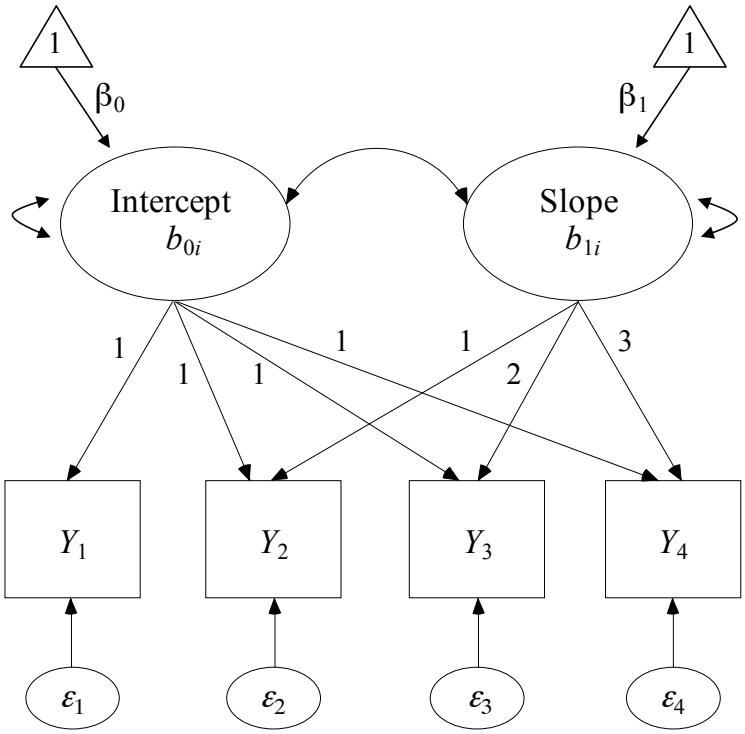


Figure 2. Path diagram of a linear Wu and Carroll growth model.

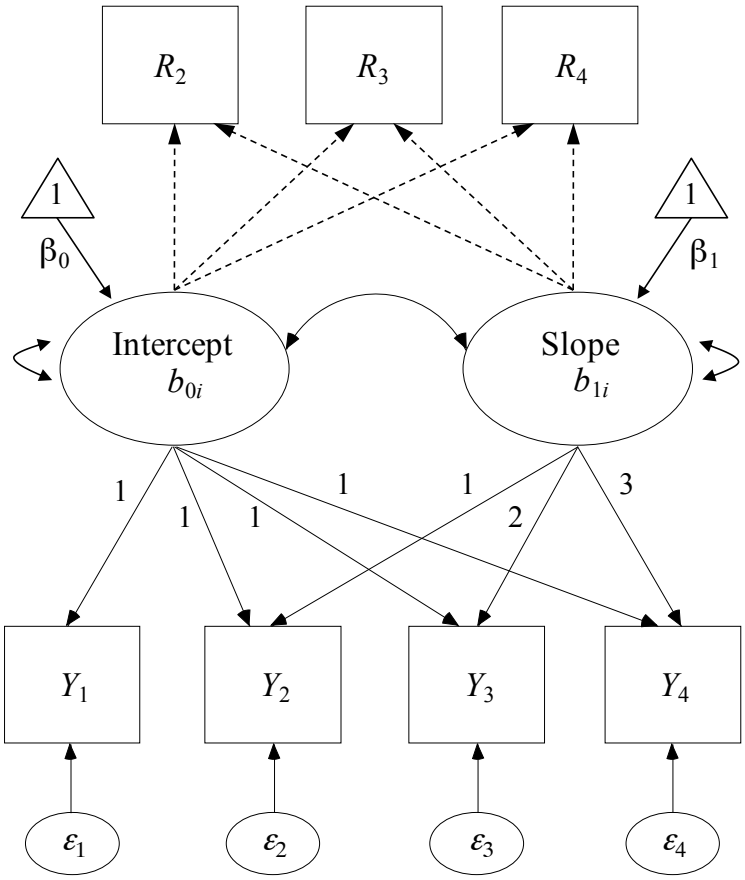


Figure 3. Path diagram of Diggle and Kenward linear growth model.

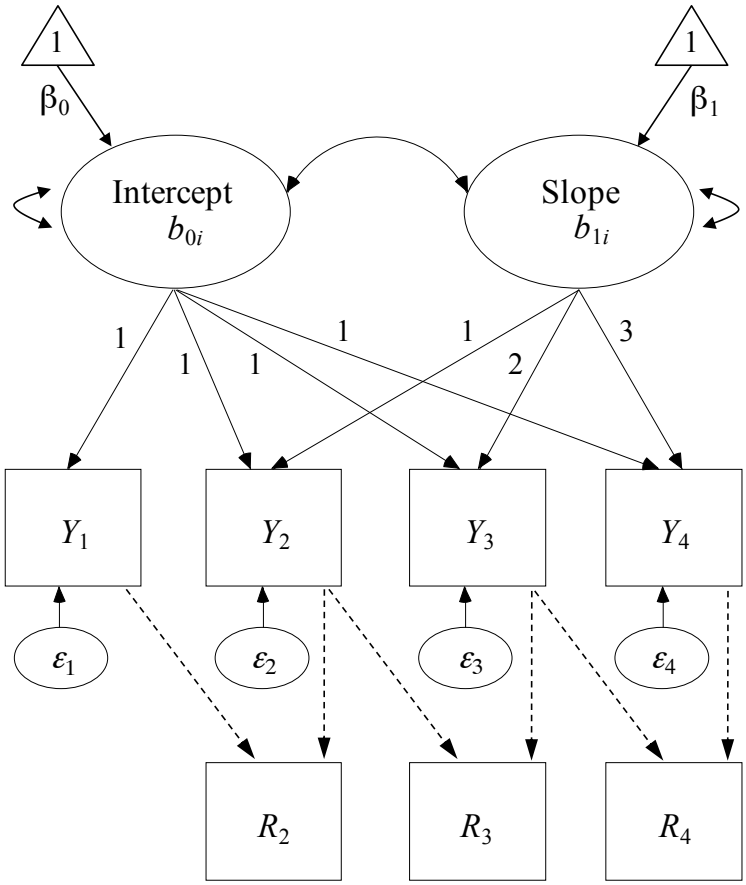


Figure 4. Observed means and fitted trajectories for each of the nine missing data patterns in the psychiatric trial data. The shaded circles denote the drug condition means and the clear circles represent the placebo group means.

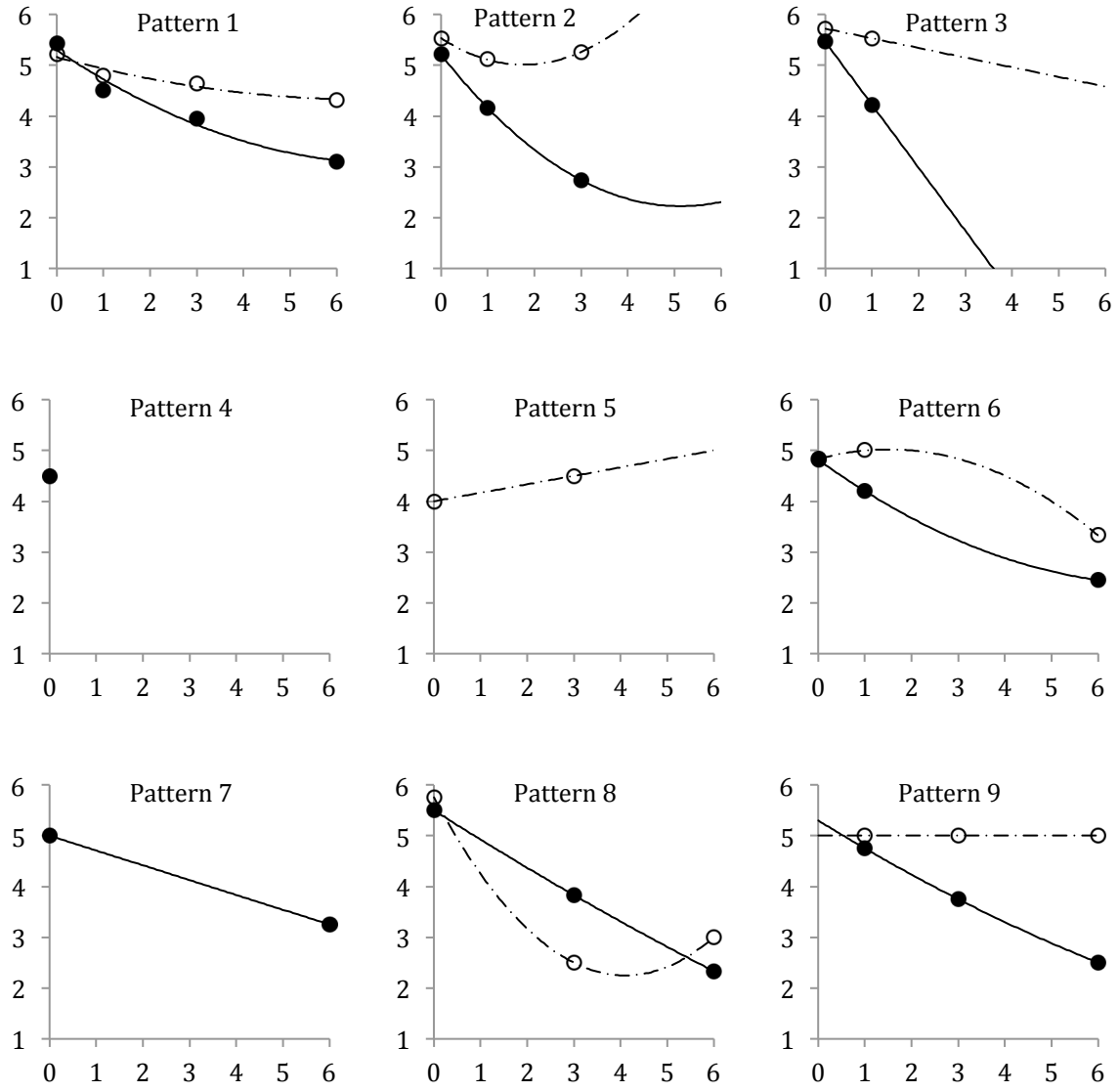


Figure 5. Quadratic growth model for the psychiatric data. Note that figure omits the latent variable intercepts and the residual covariances among the latent variables in order to reduce visual clutter.

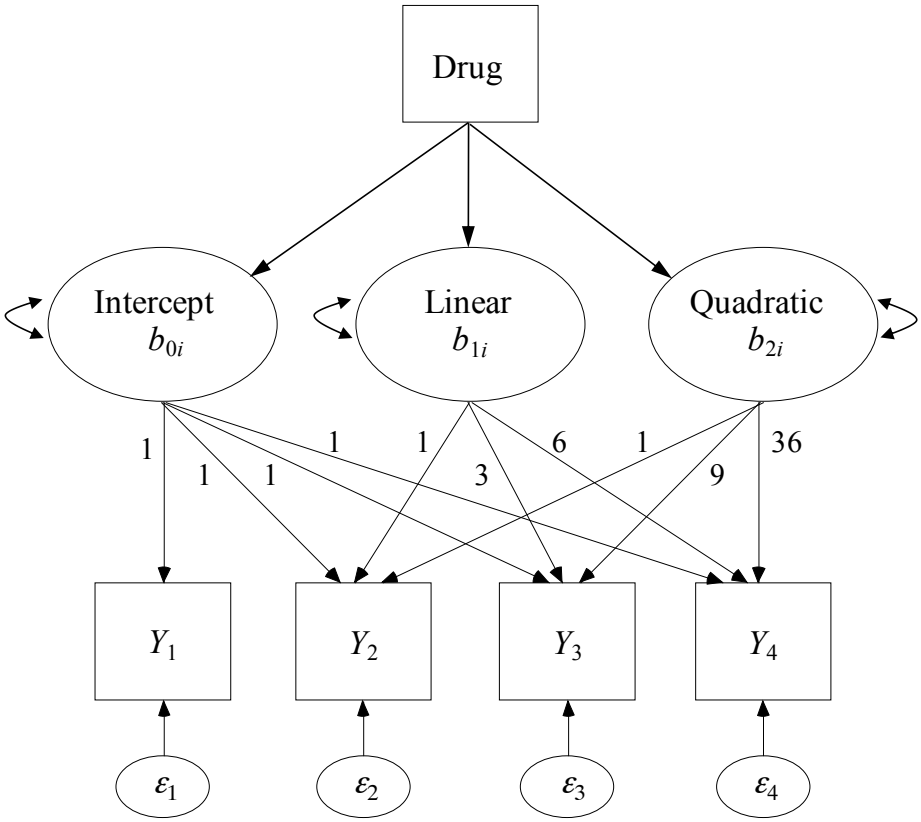


Figure 6. Model-implied growth trajectories from the MAR-based maximum likelihood analysis.

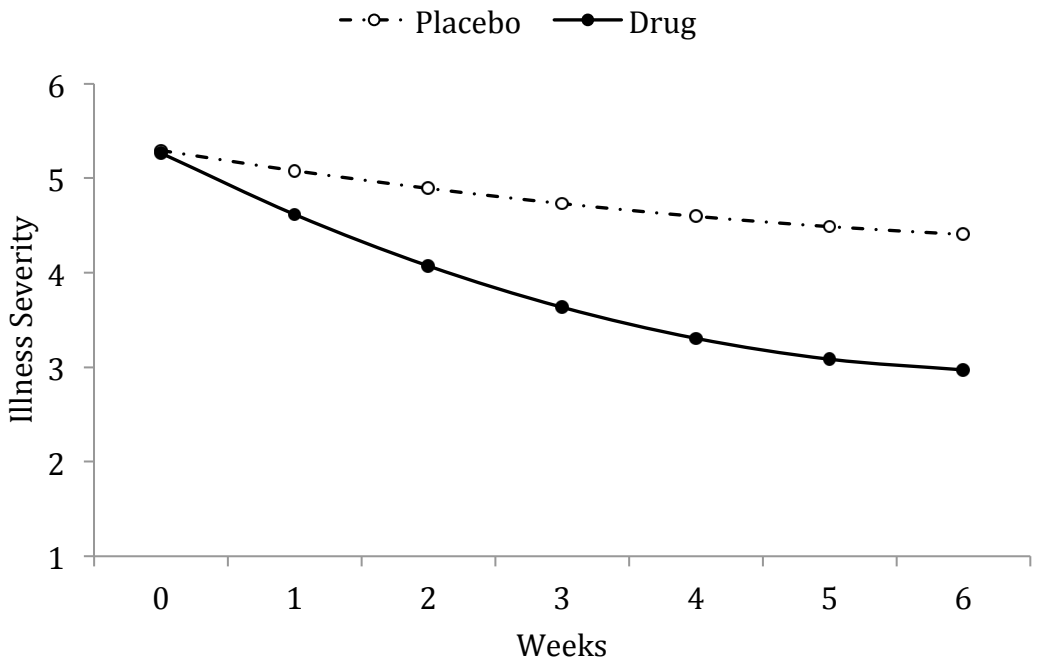


Figure 7. Diggle and Kenward quadratic growth model for the psychiatric data. Note that figure omits the latent variable intercepts and the residual covariances among the latent variables in order to reduce visual clutter.

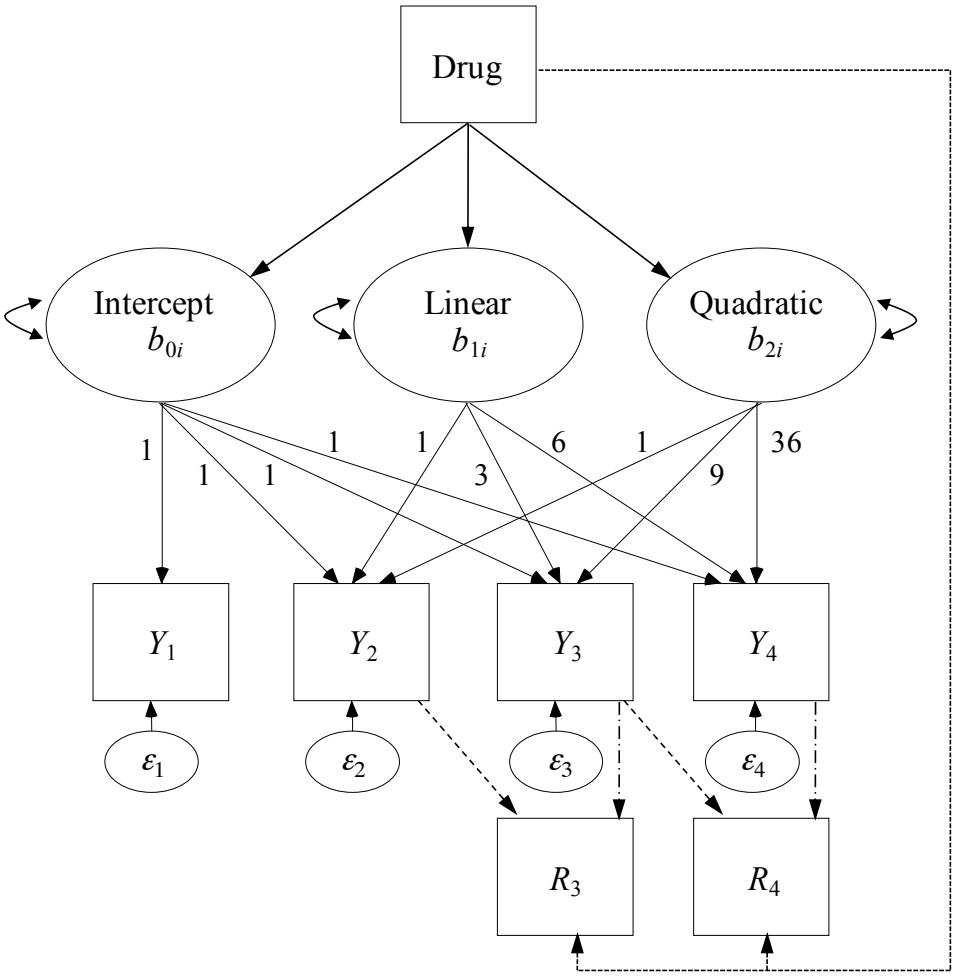


Figure 8. Wu and Carroll quadratic growth model for the psychiatric data. Note that figure omits the latent variable intercepts and the residual covariances among the latent variables in order to reduce visual clutter.

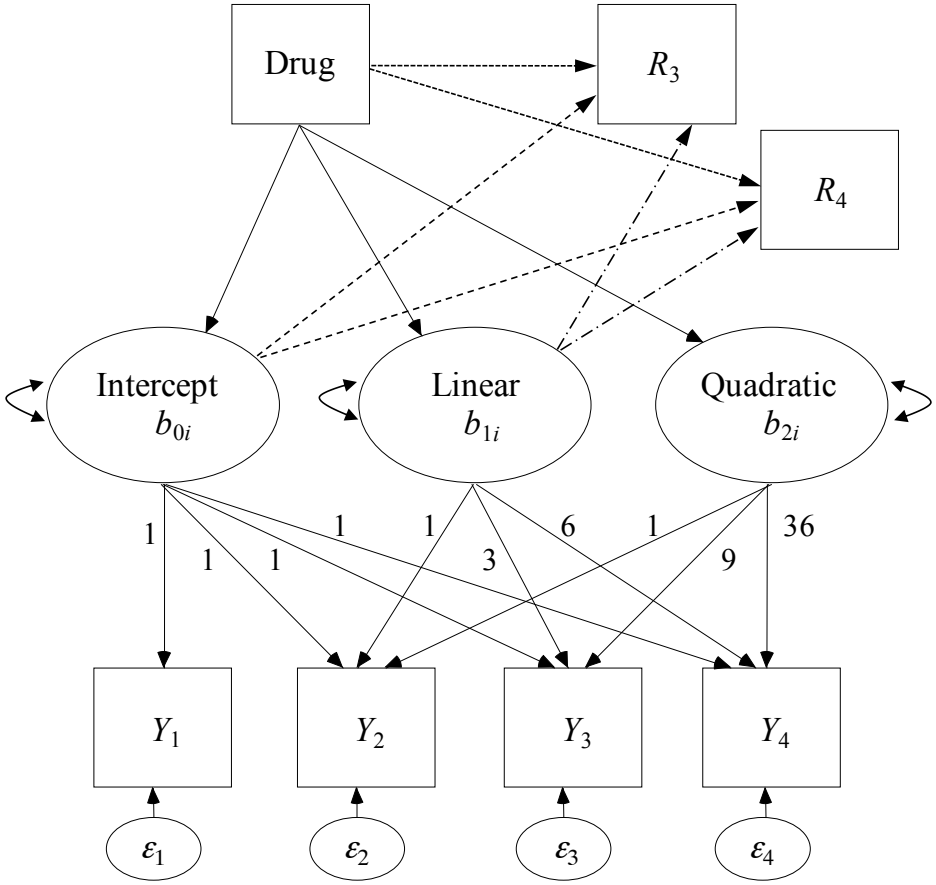
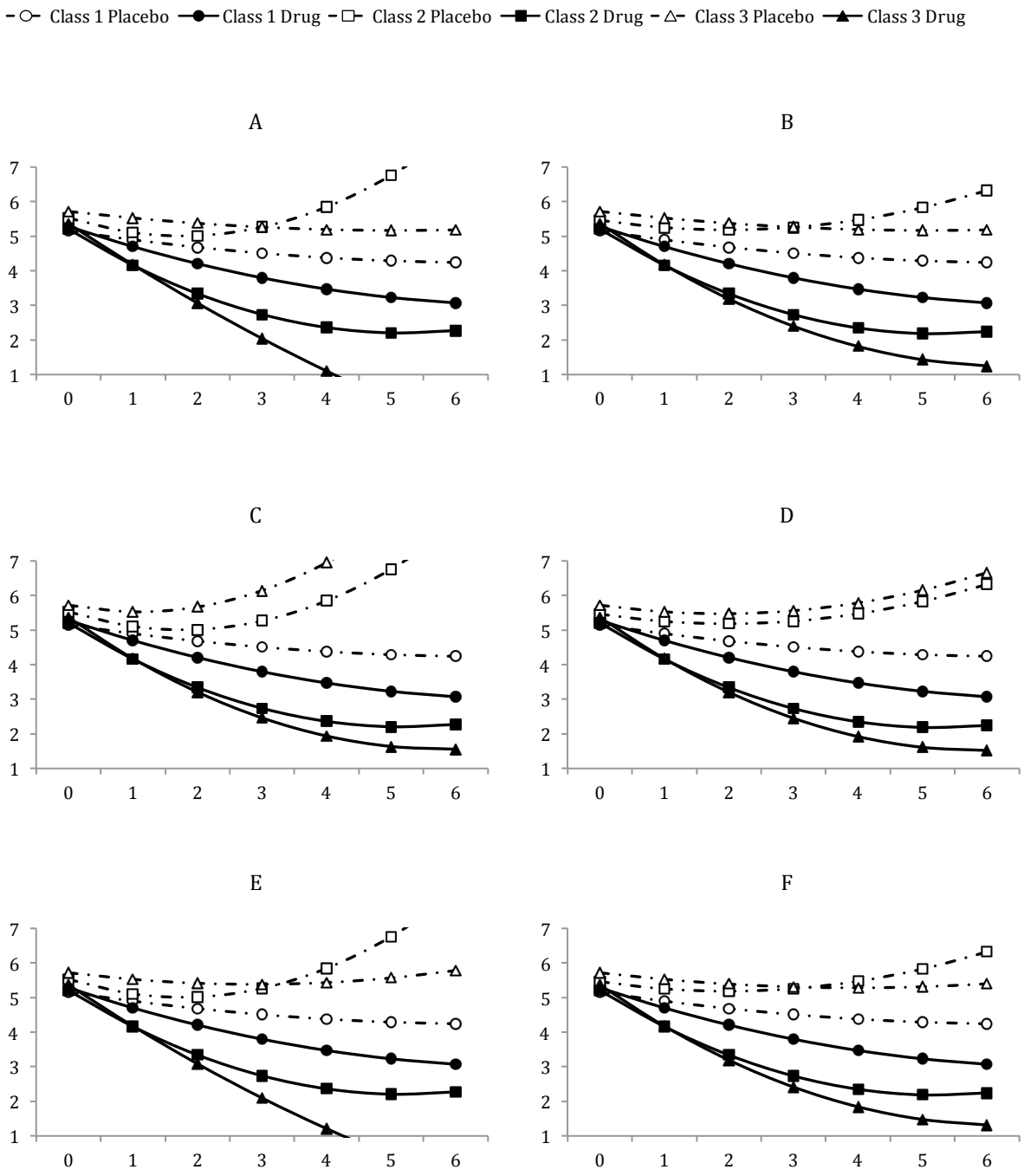


Figure 9. Class-specific model-implied growth trajectories. The complete case missing variable restriction generated panels A and B, the neighboring case missing variable restriction produced panels C and D, and the available case missing variable restriction produced E and F. Panels A, C, and E are implausible trajectories from initial analyses.



Missing Not At Random Models for Latent Growth Curve Analyses

Mplus Syntax Files

```
TITLE:

MAR-based maximum likelihood analysis;

DATA:

file is drugtrial.dat;

VARIABLE:

names are id drug male y1 y2 y3 y4 pattnum;

! id: participant id;
! drug: 0 = placebo, 1 = drug;
! male: 0 = female, 1 = male;
! y1: baseline severity score;
! y2: week 1 severity score;
! y3: week 3 severity score;
! y4: week 6 severity score;
! pattnum: missing data patterns from table 1;

usevariables are drug y1 y2 y3 y4;
missing are all (99);

MODEL:

! quadratic growth model;
! () are parameter labels used in model constraint section;

icept linear quad | y1@0 y2@1 y3@3 y4@6;
y1 - y4 (resvar);
icept (iceptvar);
icept on drug (b3);
linear on drug (b4);
quad on drug (b5);
```

```
MODEL CONSTRAINT:

! define new parameters not in the model;

new(meandiff effsize);

! compute endpoint mean difference (equation 9);

meandiff = b3 + 6*b4 + 36*b5;
effsize = meandiff / sqrt(iceptvar + resvar);

PLOT:

type is plot3;
series is y1 - y4 (linear);
```

```
TITLE:

Diggle and Kenward selection model with survival indicators;

DATA:

file is drugtrial.dat;

VARIABLE:

names are id drug male y1 y2 y3 y4 pattnum;

! id: participant id;
! drug: 0 = placebo, 1 = drug;
! male: 0 = female, 1 = male;
! y1: baseline severity score;
! y2: week 1 severity score;
! y3: week 3 severity score;
! y4: week 6 severity score;
! pattnum: missing data patterns from table 1;

usevariables are drug y1 y2 y3 y4 r3 r4;
missing are all (99);
categorical are r3 r4;

! create survival model indicators, r3 and r4;

DATA MISSING:

names = y2 y3 y4;
type = sdropout;
binary = r3 r4;

ANALYSIS:

estimator = mlr;
integration = montecarlo;
starts = 100 100;

MODEL:

! quadratic growth model;
! () are parameter labels used in model constraint section;

icept linear quad | y1@0 y2@1 y3@3 y4@6;
y1 - y4 (resvar);
icept (iceptvar);
icept on drug (b3);
linear on drug (b4);
quad on drug (b5);
```

```
! logistic regressions;

r3 on y2 (logb1);
r3 on y3 (logb2);
r3 on drug (logb3);
r4 on y3 (logb1);
r4 on y4 (logb2);
r4 on drug (logb3);

MODEL CONSTRAINT:

! define new parameters not in the model;

new(meandiff effsize);

! compute endpoint mean difference (equation 9);

meandiff = b3 + 6*b4 + 36*b5;
effsize = meandiff / sqrt(iceptvar + resvar);

PLOT:

type is plot3;
series is y1 - y4 (linear);
```

```
TITLE:

Wu and Carroll selection (shared parameter) model with survival indicators;

DATA:

file is drugtrial.dat;

VARIABLE:

names are id drug male y1 y2 y3 y4 pattnum;

! id: participant id;
! drug: 0 = placebo, 1 = drug;
! male: 0 = female, 1 = male;
! y1: baseline severity score;
! y2: week 1 severity score;
! y3: week 3 severity score;
! y4: week 6 severity score;
! pattnum: missing data patterns from table 1;

usevariables are drug y1 y2 y3 y4 r3 r4;
categorical are r3 r4;
missing are all (99);

! create survival model indicators, r3 and r4;

DATA MISSING:

names = y2 y3 y4;
type = sdropout;
binary = r3 r4;

ANALYSIS:

estimator = mlr;
integration = montecarlo;
starts = 100 100;

MODEL:

! quadratic growth model;
! () are parameter labels used in model constraint section;

icept linear quad | y1@0 y2@1 y3@3 y4@6;
y1 - y4 (resvar);
icept (iceptvar);
quad@0;
icept on drug (b3);
linear on drug (b4);
quad on drug (b5);
```

```
! logistic regressions;

r3 on icept (logb1);
r3 on linear (logb2);
r4 on icept (logb1);
r4 on linear (logb2);

MODEL CONSTRAINT:

! define new parameters not in the model;

new(meandiff effsize);

! compute endpoint mean difference (equation 9);

meandiff = b3 + 6*b4 + 36*b5;
effsize = meandiff / sqrt(iceptvar + resvar);

PLOT:

type is plot3;
series is y1 - y4 (linear);
```

```
TITLE:

pattern mixture with complete case restriction;

DATA:

file is drugtrial.dat;

VARIABLE:

names are id drug male y1 y2 y3 y4 pattnum;

! id: participant id;
! drug: 0 = placebo, 1 = drug;
! male: 0 = female, 1 = male;
! y1: baseline severity score;
! y2: week 1 severity score;
! y3: week 3 severity score;
! y4: week 6 severity score;
! pattnum: missing data patterns from table 1;

usevariables are drug y1 y2 y3 y4 group;
missing are all (99);

! specify pseudo latent class variable and number of classes;

classes = pattern(3);

! define latent classes with manifest grouping variable;

knownclass = pattern(group = 1 group = 2 group = 3);

DEFINE:

! collapse missing data patterns into grouping variable;

if (pattnum eq 1 or pattnum ge 5) then group = 1;
if (pattnum eq 2) then group = 2;
if (pattnum eq 3 or pattnum eq 4) then group = 3;

ANALYSIS:

type = mixture;

MODEL:

%overall%

! quadratic growth model;
! () are parameter labels used in model constraint section;

icept linear quad | y1@0 y2@1 y3@3 y4@6;
y1 - y4 (resvar);
icept (iceptvar);
icept on drug;
linear on drug;
quad on drug;
```

```
! latent variable means used to compute proportions;

[pattern#1] (p1logit);
[pattern#2] (p2logit);

! class-specific models;

%pattern#1%

[icept] (icept1);
[linear] (linear1);
[quad] (quad1);
icept on drug (iondrug1);
linear on drug (londrug1);
quad on drug (qondrug1);

%pattern#2%

[icept] (icept2);
[linear] (linear2);
[quad] (quad2);
icept on drug (iondrug2);
linear on drug (londrug2);
quad on drug (qondrug2);

%pattern#3%


[icept] (icept3);
[linear] (linear3);
[quad] (quad3);
icept on drug (iondrug3);
linear on drug (londrug3);
quad on drug (qondrug3);

MODEL CONSTRAINT:

! implement identifying restrictions;

quad3 = quad1;
qondrug3 = .08;
quad2 < .07;

! define new parameters not in the model;

new(pi1 pi2 pi3 b0 b1 b2 b3 b4 b5 meandiff effsize);

! compute pattern proportions;

pi1 = exp(p1logit)/(exp(0) + exp(p1logit) + exp(p2logit));
pi2 = exp(p2logit)/(exp(0) + exp(p1logit) + exp(p2logit));
pi3 = exp(0)/(exp(0) + exp(p1logit) + exp(p2logit));
```

```
! compute average estimates (equation 10);

b0 = pi1*icept1 + pi2*icept2 + pi3*icept3;
b1 = pi1*linear1 + pi2*linear2 + pi3*linear3;
b2 = pi1*quad1 + pi2*quad2 + pi3*quad3;
b3 = pi1*iondrug1 + pi2*iondrug2 + pi3*iondrug3;
b4 = pi1*londrug1 + pi2*londrug2 + pi3*londrug3;
b5 = pi1*qondrug1 + pi2*qondrug2 + pi3*qondrug3;

! compute endpoint mean difference (equation 9);

meandiff = b3 + 6*b4 + 36*b5;
effsize = meandiff / sqrt(iceptvar + resvar);

PLOT:

type is plot3;
series is y1 - y4 (linear);
```

```
TITLE:

pattern mixture with neighboring case restriction;

DATA:

file is drugtrial.dat;

VARIABLE:

names are id drug male y1 y2 y3 y4 pattnum;

! id: participant id;
! drug: 0 = placebo, 1 = drug;
! male: 0 = female, 1 = male;
! y1: baseline severity score;
! y2: week 1 severity score;
! y3: week 3 severity score;
! y4: week 6 severity score;
! pattnum: missing data patterns from table 1;

usevariables are drug y1 y2 y3 y4 group;
missing are all (99);

! specify pseudo latent class variable and number of classes;

classes = pattern(3);

! define latent classes with manifest grouping variable;

knownclass = pattern(group = 1 group = 2 group = 3);

DEFINE:

! collapse missing data patterns into grouping variable;

if (pattnum eq 1 or pattnum ge 5) then group = 1;
if (pattnum eq 2) then group = 2;
if (pattnum eq 3 or pattnum eq 4) then group = 3;

ANALYSIS:

type = mixture;

MODEL:

%overall%

! quadratic growth model;
! () are parameter labels used in model constraint section;

icept linear quad | y1@0 y2@1 y3@3 y4@6;
y1 - y4 (resvar);
icept (iceptvar);
icept on drug;
linear on drug;
quad on drug;
```

```
! latent variable means used to compute proportions;

[pattern#1] (p1logit);
[pattern#2] (p2logit);

! class-specific models;

%pattern#1%

[icept] (icept1);
[linear] (linear1);
[quad] (quad1);
icept on drug (iondrug1);
linear on drug (londrug1);
quad on drug (qondrug1);

%pattern#2%

[icept] (icept2);
[linear] (linear2);
[quad] (quad2);
icept on drug (iondrug2);
linear on drug (londrug2);
quad on drug (qondrug2);

%pattern#3%

[icept] (icept3);
[linear] (linear3);
[quad] (quad3);
icept on drug (iondrug3);
linear on drug (londrug3);
quad on drug (qondrug3);

MODEL CONSTRAINT:

! implement identifying restrictions;

quad3 = quad2;
qondrug3 = qondrug2;
quad2 < .07;

! define new parameters not in the model;

new(pi1 pi2 pi3 b0 b1 b2 b3 b4 b5 meandiff effsize);

! compute pattern proportions;

pi1 = exp(p1logit)/(exp(0) + exp(p1logit) + exp(p2logit));
pi2 = exp(p2logit)/(exp(0) + exp(p1logit) + exp(p2logit));
pi3 = exp(0)/(exp(0) + exp(p1logit) + exp(p2logit));
```

```
! compute average estimates (equation 10);

b0 = pi1*icept1 + pi2*icept2 + pi3*icept3;
b1 = pi1*linear1 + pi2*linear2 + pi3*linear3;
b2 = pi1*quad1 + pi2*quad2 + pi3*quad3;
b3 = pi1*iondrug1 + pi2*iondrug2 + pi3*iondrug3;
b4 = pi1*londrug1 + pi2*londrug2 + pi3*londrug3;
b5 = pi1*qondrug1 + pi2*qondrug2 + pi3*qondrug3;

! compute endpoint mean difference (equation 9);

meandiff = b3 + 6*b4 + 36*b5;
effsize = meandiff / sqrt(iceptvar + resvar);

PLOT:

type is plot3;
series is y1 - y4 (linear);
```

```
TITLE:

pattern mixture with available case restriction;

DATA:

file is drugtrial.dat;

VARIABLE:

names are id drug male y1 y2 y3 y4 pattnum;

! id: participant id;
! drug: 0 = placebo, 1 = drug;
! male: 0 = female, 1 = male;
! y1: baseline severity score;
! y2: week 1 severity score;
! y3: week 3 severity score;
! y4: week 6 severity score;
! pattnum: missing data patterns from table 1;

usevariables are drug y1 y2 y3 y4 group;
missing are all (99);

! specify pseudo latent class variable and number of classes;

classes = pattern(3);

! define latent classes with manifest grouping variable;

knownclass = pattern(group = 1 group = 2 group = 3);

DEFINE:

! collapse missing data patterns into grouping variable;

if (pattnum eq 1 or pattnum ge 5) then group = 1;
if (pattnum eq 2) then group = 2;
if (pattnum eq 3 or pattnum eq 4) then group = 3;

ANALYSIS:

type = mixture;

MODEL:

%overall%

! quadratic growth model;
! () are parameter labels used in model constraint section;

icept linear quad | y1@0 y2@1 y3@3 y4@6;
y1 - y4 (resvar);
icept (iceptvar);
icept on drug;
linear on drug;
quad on drug;
```

```
! latent variable means used to compute proportions;

[pattern#1] (p1logit);
[pattern#2] (p2logit);

! class-specific models;

%pattern#1%

[icept] (icept1);
[linear] (linear1);
[quad] (quad1);
icept on drug (iondrug1);
linear on drug (londrug1);
quad on drug (qondrug1);

%pattern#2%

[icept] (icept2);
[linear] (linear2);
[quad] (quad2);
icept on drug (iondrug2);
linear on drug (londrug2);
quad on drug (qondrug2);

%pattern#3%

[icept] (icept3);
[linear] (linear3);
[quad] (quad3);
icept on drug (iondrug3);
linear on drug (londrug3);
quad on drug (qondrug3);

MODEL CONSTRAINT:

! implement identifying restrictions;

quad2 < .07;
quad3 = (336/389)*quad1 + (53/389)*quad2;
qondrug3 = (336/389)*.08 + (53/389)*qondrug2;


! define new parameters not in the model;

new(pi1 pi2 pi3 b0 b1 b2 b3 b4 b5 meandiff effsize);

! compute pattern proportions;

pi1 = exp(p1logit)/(exp(0) + exp(p1logit) + exp(p2logit));
pi2 = exp(p2logit)/(exp(0) + exp(p1logit) + exp(p2logit));
pi3 = exp(0)/(exp(0) + exp(p1logit) + exp(p2logit));
```

```
! compute average estimates (equation 10);

b0 = pi1*icept1 + pi2*icept2 + pi3*icept3;
b1 = pi1*linear1 + pi2*linear2 + pi3*linear3;
b2 = pi1*quad1 + pi2*quad2 + pi3*quad3;
b3 = pi1*iondrug1 + pi2*iondrug2 + pi3*iondrug3;
b4 = pi1*londrug1 + pi2*londrug2 + pi3*londrug3;
b5 = pi1*qondrug1 + pi2*qondrug2 + pi3*qondrug3;

! compute endpoint mean difference (equation 9);

meandiff = b3 + 6*b4 + 36*b5;
effsize = meandiff / sqrt(iceptvar + resvar);

PLOT:

type is plot3;
series is y1 - y4 (linear);
```