Rehabilitation Psychology

Analyzing Longitudinal Data With Missing Values

Craig K. Enders

Online First Publication, October 3, 2011. doi: 10.1037/a0025579

Rehabilitation Psychology

0090-5550/11/$12.00 DOI: 10.1037/a0025579

Analyzing Longitudinal Data With Missing Values

Craig K. Enders
Arizona State University

Missing data methodology has improved dramatically in recent years, and popular computer programs now offer a variety of sophisticated options. Despite the widespread availability of theoretically justified methods, researchers in many disciplines still rely on subpar strategies that either eliminate incomplete cases or impute the missing scores with a single set of replacement values. This article provides readers with a nontechnical overview of some key issues from the missing data literature and demonstrates several of the techniques that methodologists currently recommend. This article begins by describing Rubin's missing data mechanisms. After a brief discussion of popular ad hoc approaches, the article provides a more detailed description of five analytic approaches that have received considerable attention in the missing data literature: maximum likelihood estimation, multiple imputation, the selection model, the shared parameter model, and the pattern mixture model. Finally, a series of data analysis examples illustrate the application of these methods.

Keywords: missing data, maximum likelihood estimation, multiple imputation, longitudinal analyses, multilevel model

Missing data are one of the most common analytic problems in the behavioral sciences. For decades, researchers were forced to rely on a variety of subpar strategies that either eliminated incomplete cases or imputed the missing scores with a single set of replacement values. These older techniques are problematic because they require strict assumptions about the cause of missing data and are prone to bias in most situations. The origin of modern missing data handling techniques dates back to the 1970s, when Rubin (1976) developed a theoretical framework for missing data problems and methodologists developed the underpinnings of two approaches that many researchers now regard as the state of the art, maximum likelihood estimation and multiple imputation (Schafer & Graham, 2002). Since then, the body of missing data literature has grown steadily, and most popular computer programs now offer a variety of sophisticated analytic options.

Despite the fact that they have fallen out of favor in the methodological literature (e.g., Wilkinson & Task Force on Statistical Inference, 1999), older ad hoc missing data handling approaches continue to enjoy widespread use in published research articles (Bodner, 2006; Peugh & Enders, 2004). Anecdotally, my experience suggests that a number of "urban myths" are partially responsible for the continued reliance on older approaches (e.g., all missing data techniques are created equal; multiple imputation is on par with making up data; sophisticated missing data handling techniques only work when the percentage of missing values is small; modern techniques are only appropriate when the missing values are isolated to the dependent variable). Additionally, some researchers have the misperception that sophisticated missing data handling techniques are difficult to implement. Although this used to be the case, it is no longer true; doing the right thing is usually just as easy as doing the wrong thing.

Given that researchers still routinely apply approaches that are inconsistent with methodological best practice, the purpose of this article is to provide readers with a nontechnical overview of some key issues from the missing data literature. I begin by describing Rubin's (1976) theoretical framework because his so-called missing data mechanisms provide a basis for understanding when and why different missing data handling techniques work or fail. Next, I provide a brief description of several analytic approaches, with a particular emphasis on five methods that have received considerable attention in the missing data literature: maximum likelihood estimation, multiple imputation, the selection model, the shared parameter model, and the pattern mixture model. Finally, I present a series of data analysis examples that illustrate the application of these methods.

Motivating Example

Throughout the article, I use a longitudinal study of depression to illustrate various concepts. Although I could have used a real data set for this purpose, I chose to use artificial data because this allows us to evaluate the performance of different techniques in the subsequent analysis examples (i.e., I generated the data set to produce certain parameter values prior to imposing missing data). The artificial data set mimics a randomized depression trial where researchers assign $n = 280$ participants to either a treatment or a control condition. The researchers then collect depression scores at baseline, at a 1-month follow-up, and at a 2-month follow-up. Table 1 gives the complete-data descriptive statistics for both groups.

To mimic a realistic dropout mechanism, I imposed missing values such that 13.6% of the participants dropped out after the baseline assessment and an additional 15.0% of the cases dropped out after the 1-month follow-up. Table 2 shows the resulting

Correspondence concerning this article should be addressed to Craig K. Enders, PhD, Department of Psychology, Arizona State University, Box 871104, Tempe, AZ 85287-1104. E-mail: craig.enders@asu.edu

Table 1
Complete-Data Descriptive Statistics From the Depression Data Set

Assessment	*M*	*SD*
Treatment (n = 141)		
Baseline	35.99	9.49
One-month	35.03	11.42
Two-month	35.60	12.71
Control (n = 139)		
Baseline	34.95	10.43
One-month	31.35	11.77
Two-month	30.07	11.47

missing data patterns (the missing data literature refers to this configuration of missing values as a monotone missing data pattern). Furthermore, the missing values represent a mixture of dropout mechanisms. For the participants that dropped out after the baseline assessment (i.e., pattern 3), the probability of missing data was either related to baseline scores or to the would-be values from the 1-month follow-up. For the participants that dropped out before the final assessment (i.e., pattern 2), the probability of attrition was always dependent on the would-be score from the final wave. The dropout mechanism also differed between the intervention groups, such that treatment cases in the lower range of a given score distribution had a higher probability of dropout, whereas control cases in the upper range of a particular score distribution were more likely to leave the study. Among other things, this deletion process mimics a situation where treatment cases that experience mild symptoms or rapid improvement have a higher tendency to quit the study, whereas control cases that experience severe symptoms or no improvement have a higher rate of attrition.

Review of the Multilevel Growth Model

Kwok et al.'s (2008) tutorial on multilevel models is one of the more highly cited papers from *Rehabilitation Psychology* in the past five years. Because of the growing interest in this technique, I use a longitudinal multilevel model (also known as a growth model, hierarchical linear model, mixed linear model, and latent curve model) as a springboard for describing different missing data handling techniques. Researchers typically use either the multilevel or the structural equation modeling framework to estimate growth models. This section gives a brief description of both approaches, and Kwok et al. (2008) and others (e.g., Bollen & Curran, 2006; Raudenbush & Bryk, 2002; Singer & Willett, 2003) provide additional details. To avoid getting bogged down by the idiosyncratic notational system of a particular modeling paradigm, I use a generic notation scheme that expresses the growth model as a linear regression equation.

The linear growth model expresses the outcome variable as a function of a temporal predictor variable that captures the passage of time (e.g., weeks or months since the baseline measurement; weeks or months before the final assessment). The model is

$$Y_{ti} = \beta_0 + \beta_1(TIME_{ti}) + b_{0i} + b_{1i}(TIME_{ti}) + e_{ti} \quad (1)$$

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, β_0 is the intercept, and β_1 is a slope coefficient that quantifies the expected change in the outcome variable for a one-unit increment in the *TIME* variable (e.g., the average change per month). Growth models are powerful tools for assessing change because they allow researchers to estimate the average developmental trend (i.e., the linear trend defined by β_0 and β_1) as well as individual heterogeneity around this trajectory. To this end, b_{0i} and b_{1i} are residuals that allow the intercepts and the slopes to vary across individuals. Finally, e_{ti} is a time-specific residual that captures discrepancies between the idealized linear trajectories and the observed scores. Note that a growth curve analysis estimates the variance of each residual as opposed to the residuals themselves. For example, the variance of b_{1i} quantifies the degree to which the individual change trajectories deviate around the average change rate.

Before going further, the coding of the *TIME* variable in Equation 1 warrants a brief explanation. To facilitate the interpretation of the intercept coefficient, researchers typically express *TIME* relative to some fixed point. For example, recall that the depression data set contains a baseline assessment, scores from a 1-month follow-up, and scores from a 2-month follow-up. For reasons that I explain later, I centered the temporal predictor relative to the first assessment, such that *TIME* equaled 0 at baseline, 1 at the 1-month follow-up, and 2 at the 2-month follow-up. Under this coding scheme, β_0 represents the baseline depression average; in a linear model, β_1 is unaffected by centering and still represents the average monthly change.

The growth model can readily incorporate predictor variables that explain individual differences in the developmental trajectories. For example, consider the preceding depression example where researchers assign participants to one of two treatment arms. In this case, the model becomes

$$Y_{ti} = \beta_0 + \beta_1(TIME_{ti}) + \beta_2(TX_i) + \beta_3(TIME_{ti})(TX_i) + b_{0i} + b_{1i}(TIME_{ti}) + e_{ti} \quad (2)$$

where TX_i is a dummy variable that denotes treatment group membership (i.e., 0 = control, 1 = treatment). The interpretation of the regression coefficients (i.e., the fixed effects) follows standard linear regression, such that β_0 and β_1 represent the baseline average and the growth rate for the control group, respectively, β_2 denotes the baseline mean difference between the two groups, and β_3 quantifies the slope difference (i.e., the group by time interaction).

Cast as a structural equation model, Equation 2 is a two-factor confirmatory factor analysis with a mean structure. To illustrate, Figure 1 depicts the model as a path diagram, where ellipses denote latent variables, rectangles symbolize measured variables, single-headed straight arrows represent regression coefficients, and

Table 2
Missing Data Patterns From the Depression Data Set

	Data collection wave			
Pattern	Baseline	One-month	Two-month	% of sample
1	O	O	O	71.4%
2	O	O	M	15.0%
3	O	M	M	13.6%

Note. O = observed; M = missing.

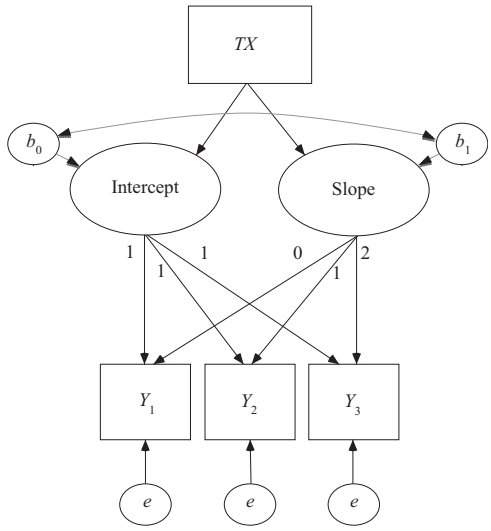

Figure 1. Path diagram of a linear growth model with a binary treatment group indicator predicting individual intercepts and slopes.

the double-headed curved arrow is a covariance. The factor model appears very different from Equation 2, but it is equivalent and yields identical estimates, at least in this example. Specifically, the latent factors capture the individual intercepts and slopes (i.e., the b_{0i} and b_{1i} terms), and the corresponding latent variable means quantify the control group means from Equation 2 (i.e., β_0 and β_1).[1] Further, the path coefficients that connect the treatment indicator to the latent factors correspond with the mean difference parameters from the multilevel model (i.e., β_2 and β_3). Finally, the pattern of factor loadings specifies the functional form of the developmental trajectory. Specifically, the unit factor loadings for the intercept latent factor reflect the fact that the baseline is a constant component of each predicted score, and the slope factor loadings represent the values of the *TIME* variable.

Although there is a one-to-one equivalence between the multilevel model and the structural equation model, software programs that implement these models differ rather dramatically in their estimation capabilities. Multilevel modeling software packages (e.g., the MIXED procedures in SAS or SPSS) can implement some of the techniques that I describe in the article, provided that the missing values are isolated to the repeated measures variables. Structural equation modeling programs are more flexible because they can accommodate missing data on any variable in the model (see Enders, 2010, pp. 116–118, for a discussion of missing explanatory variables). Consequently, I use the structural equation modeling framework in the subsequent analysis examples. In particular, I use Mplus (Muthén & Muthén, 1998-2010) because it can accommodate categorical and continuous outcome variables in the same analysis (this functionality is necessary for two of the missing data models that I describe in a later section).

Missing Data Mechanisms

Rubin (1976) proposed a classification system for missing data problems that is now firmly entrenched in the methodological literature. In Rubin's paradigm, each participant has a score on a particular variable (the score may or may not be observed) and a probability of a missing value on that variable. The nature of the association between the probability of missing data and other variables defines one of three so-called missing data mechanisms. For example, in the depression data set, the propensity for dropout at the final wave may be related to predictor variables such as treatment group membership, to depression scores from previous assessments, or to the would-be scores from that assessment; although researchers usually discount the possibility, the probability of missing data may be completely unrelated to other variables. From a practical perspective, these different scenarios effectively serve as assumptions for a missing data analysis. As I explain later, ad hoc missing data techniques (e.g., deleting incomplete cases) tend to assume that dropout is unrelated to the data, whereas modern approaches work from more lenient assumptions.

A missing completely at random (MCAR) mechanism occurs when the propensity for missing data on a particular variable is unrelated to other measured variables and to the would-be values of that variable. Returning to the depression data, MCAR requires that the propensity for a missing depression score at wave *t* is unrelated to (a) treatment group status, (b) depression scores from previous assessments, and (c) the would-be depression scores from wave *t*. MCAR is the most benign (and perhaps the most unrealistic) of Rubin's mechanisms because the cases with missing data are no different from the cases with complete data, on average.

A missing at random (MAR) mechanism holds when the probability of missing data on a variable is related to other variables, but not to the would-be values of the incomplete variable. Returning to the depression data, MAR allows the propensity for missing data at wave *t* to relate to treatment group status or to depression scores from previous assessments. The important stipulation is that, controlling for treatment group status and previous scores, the would-be values from wave *t* have no association with the likelihood of dropout. Notice that MAR is a less stringent assumption than MCAR because it accommodates systematic missingness.

Finally, a not missing at random (NMAR) mechanism occurs when the probability of missing data on a variable is related to the would-be value of that variable (i.e., outcome-dependent missingness). Reconsidering the depression data, NMAR implies the likelihood of dropout at wave *t* is associated with the would-be depression score from that assessment, even after controlling for treatment group status and previous depression scores. Of the three mechanisms, NMAR is arguably the most problematic. Methodologists have devoted considerable energy to this problem, but NMAR-based analysis models seem to perform well in a relatively narrow range of circumstances.

Rubin's mechanisms are important to consider because they largely dictate the performance of a missing data handling technique. For example, excluding incomplete cases from an analysis (i.e., listwise deletion or complete-case analysis) is defensible with an MCAR mechanism because the complete cases are a representative sample of the hypothetically complete data set. However, this approach can introduce substantial bias under an MAR or NMAR mechanism. In contrast, maximum likelihood estimation and multiple imputation yield accurate estimates with an MCAR or MAR mechanism, but

[1] Technically, the latent variable means are regression intercepts because the latent factors are endogenous variables.

these approaches also produce biased estimates under an NMAR mechanism. Methodologists have developed a variety of NMAR analysis models for longitudinal data, three of which I describe later in the article (selection models, shared parameter models, and pattern mixture models). However, implementing these techniques is difficult because the analysis must incorporate an additional model that explains the probability of missing data. For example, a selection model augments the growth curve model in Equation 2 with a set of logistic regression equations that predict dropout from the repeated measures variables. This missing data model generally requires strict and untestable assumptions that go beyond the missing data mechanism, and violating these assumptions can, again, introduce substantial bias.

It is safe to say that the methodological literature offers little support for MCAR-based missing data handling approaches (e.g., deletion) because these methods are prone to bias in most realistic situations; dozens of published computer simulation studies have demonstrated this point. Unfortunately, there is no way to determine whether an MAR- or NMAR-based analysis is appropriate because both mechanisms involve propositions about the *unobserved* (i.e., would-be) score values. In my experience, researchers in some disciplines are often quick to discount MAR-based analyses on grounds that NMAR models allow for outcome-dependent missingness. However, the fact that NMAR models require strict assumptions that limit their practical utility has led some methodologists to argue that MAR-based analysis are often more defensible (Demirtas & Schafer, 2003; Enders & Gottschall, 2011; Schafer, 2003). Ultimately, researchers need to construct logical arguments that defend their analytic choices because the observed data cannot inform model selection. Given the difficulty of defending a set of untestable assumptions, performing a sensitivity analysis that fits MAR- and NMAR-based models to the same data is often a sensible strategy. I illustrate this approach later in the article.

Ad Hoc Missing Data Methods

During the past 50 years, literally dozens of ad hoc missing techniques have appeared in the literature (I refer to them as ad hoc because they predate Rubin's seminal work and thus have no theoretical justification). Generally speaking, these approaches either eliminate incomplete cases or impute the missing scores with a single set of replacement values (i.e., single imputation). Unfortunately, neither strategy tends to work well. I provide a brief description of a few common ad hoc techniques, and more detailed descriptions are available elsewhere in the literature (e.g., Enders, 2010; Schafer & Graham, 2002).

Deletion methods have enjoyed widespread use in the behavioral sciences (Peugh & Enders, 2004), perhaps because they are the default approaches in general use statistical packages such as SPSS and SAS. Listwise deletion completely eliminates cases with missing data, whereas pairwise deletion discards cases on an analysis-by-analysis basis. A decrease in power is an obvious consequence of eliminating data, but deletion approaches also assume the rather strict MCAR mechanism (i.e., the propensity for missing data is unrelated to other variables). Computer simulation studies have repeatedly demonstrated that eliminating data introduces substantial bias when the mechanism is MAR or NMAR (see Enders, 2010). For this reason, the APA Task Force on Statistical Inference (Wilkinson and Task Force on Statistical Inference, 1999) strongly discouraged the use of deletion approaches, stating that these methods are "among the worst methods available for practical applications" (p. 598).

A number of ad hoc techniques fill in the missing values with a single set of replacement scores—this strategy contrasts with multiple imputation, which imputes missing scores with several plausible replacement values. Single imputation methods are convenient because they produce a complete data set, but they tend to produce bias regardless of the missing data mechanism, and they always attenuate standard errors. One of the oldest techniques, mean substitution, replaces missing scores with the arithmetic mean of the complete cases. Regression imputation, a historical precursor to modern MAR-based approaches, predicts the incomplete variables from the complete variables and replaces missing scores with predicted values from a regression equation. Last observation carried forward is a technique that is specific to longitudinal designs. The procedure replaces each missing value with the observed score from the preceding assessment. For example, in the depression data set, scores from the second wave would replace the missing values at the final assessment. Similarly, baseline scores would carry forward and serve as replacement values for participants with missing data at the last two waves. Although this strategy has enjoyed widespread use in the medical and clinical trials literature (Wood, White, & Thompson, 2004), methodological studies have demonstrated that it is capable of producing substantial bias, even under an MCAR mechanism (e.g., Liu & Gould, 2002; Molenberghs et al., 2004).

MAR Analysis Methods

This section outlines the two principal MAR-based analysis methods, maximum likelihood estimation and multiple imputation. These approaches have a strong theoretical foundation as well as a large body of empirical research that supports their use. Because these procedures require a less stringent assumption about the missing data mechanism (i.e., the propensity for missing data is related to other variables), they will virtually always outperform the ad hoc methods from the previous section, both with respect to accuracy and power (e.g., see Enders, 2001; Enders & Bandalos, 2001). Although multiple imputation and maximum likelihood are not yet the predominant methods in published research articles, there has been a noticeable shift to these approaches in recent years.

Maximum Likelihood Estimation

Maximum likelihood estimation identifies the population parameter values that have the highest probability of producing the sample data. Importantly, maximum likelihood uses all of the available data to generate parameter estimates; the estimator does not discard incomplete cases, nor does it impute missing values. At a conceptual level, maximum likelihood is comparable to ordinary least squares in the sense that it identifies the parameter estimates that minimize the sum of the squared distances to the observed data. For each case, the estimator uses a mathematical function called log likelihood to quantify the standardized distance between the data points and the parameters. Assuming a multivariate normal population, the log likelihood for case i is

$$\log L_i = -\frac{k_i}{2}\log(2\pi) - \frac{1}{2}\log|\Sigma_i| - \frac{1}{2}(\mathrm{Y}_i - \mu_i)^T \Sigma_i^{-1}(\mathrm{Y}_i - \mu_i) \tag{3}$$

where k_i is the number of observed scores for that individual, Y_i is vector of observed scores, and μ_i and Σ_i are estimates of the population mean vector and covariance matrix, respectively, at a particular computational cycle (in the context of a growth curve analysis, μ_i and Σ_i are model-implied matrices). In words, Equation 3 quantifies the relative probability of obtaining the values in Y_i from a multivariate normal population with a particular mean vector and covariance matrix.

Although Equation 3 is relatively complex, a small (and familiar) kernel largely drives the estimation process.

$$(\mathrm{Y}_i - \mu_i)^T \Sigma_i^{-1}(\mathrm{Y}_i - \mu_i) \tag{4}$$

Equation 4 – also known as Mahalanobis distance—is a squared z score that quantifies the standardized distance between an individual's observed scores and the parameter estimates. A small z score, and thus a high probability or high log likelihood value, results when an individual's score values are close to the variable means, whereas a large z score, and thus a low probability or low log likelihood value, results when a set of score values is distant from the means.

Like ordinary least squares estimation, maximum likelihood identifies the parameter estimates that minimize the sum of the squared distances to the data (i.e., sum of the squared z scores). To do so, it uses an aggregate log likelihood value that sums across the entire sample, as follows.

$$\log L = \Sigma \log L_i \tag{5}$$

The sample log likelihood functions much like a loss function in ordinary least squares estimation, although it is scaled such that higher values reflect a better fit to the observed data (i.e., a high log likelihood occurs when the sum of the squared z scores is small). In most situations, estimation uses an iterative optimization algorithm that repeatedly auditions different parameter values (i.e., μ and Σ or the parameters that define their model-implied counterparts) until it locates the estimates that maximize Equation 5.

Thus far, I have yet to describe how maximum likelihood accommodates missing data. Returning to Equation 3, notice that the data vector and the parameter vectors have an *i* subscript. This subscript implies that the size and the contents of the arrays can vary across individuals with different patterns of missing data. That is, the individual log likelihood equation includes the scores and parameters for which there are data and excludes the scores and parameters for which there is no data. To illustrate, reconsider the depression data set. For the cases with complete data, the squared z score computations use all three depression scores and the entire set of parameter estimates, as follows:

$$z_i^2 = \left(\begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \end{bmatrix} - \begin{bmatrix} \mu_{Y_1} \\ \mu_{Y_2} \\ \mu_{Y_3} \end{bmatrix}\right)^T \begin{bmatrix} \sigma^2_{Y_1} & \sigma_{Y_1Y_2} & \sigma_{Y_1Y_3} \\ \sigma_{Y_2Y_1} & \sigma^2_{Y_2} & \sigma_{Y_2Y_3} \\ \sigma_{Y_3Y_1} & \sigma_{Y_3Y_2} & \sigma^2_{Y_3} \end{bmatrix}^{-1} \left(\begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \end{bmatrix} - \begin{bmatrix} \mu_{Y_1} \\ \mu_{Y_2} \\ \mu_{Y_3} \end{bmatrix}\right)$$

For the participants that drop out before the final wave, the z score computations use the available data and the corresponding estimates, as follows:

$$z_i^2 = \left(\begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} - \begin{bmatrix} \mu_{Y_1} \\ \mu_{Y_2} \end{bmatrix}\right)^T \begin{bmatrix} \sigma^2_{Y_1} & \sigma_{Y_1Y_2} \\ \sigma_{Y_2Y_1} & \sigma^2_{Y_2} \end{bmatrix}^{-1} \left(\begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} - \begin{bmatrix} \mu_{Y_1} \\ \mu_{Y_2} \end{bmatrix}\right)$$

Finally, the computations simplify even further for the individuals that dropout after the baseline assessment.

$$z_i^2 = \frac{(Y_1 - \mu_{Y_1})^2}{\sigma^2_{Y_1}}$$

The previous equations illustrate how the log likelihood computations use all of the available data, but they do not explain why including the partial data records improves the accuracy of the resulting estimates. Although the estimation process does not literally impute the missing values, it does borrow information from the observed scores when estimating parameters from incomplete data. The normal distribution is integral to this process. For example, consider an individual who has relatively high depression scores at the first two assessments and a missing value at wave 3. Given that the high scores at the first two assessments originated from a multivariate normal distribution, not all wave 3 scores are equally likely. In a multivariate normal distribution, the missing score would most likely fall in the upper tail of the distribution, and it would be relatively unlikely for this individual to have a low score at the final wave. Similarly, consider an individual who scored near the center of the distribution at the first two assessments. In a multivariate normal distribution, it would be relatively unlikely for that individual to have an extreme wave 3 score in either tail of the distribution. Rather, the missing value would likely be close to the center of the distribution. In the previous examples, the normal distribution effectively constrains the range of plausible values for the missing scores. Thus, although maximum likelihood estimation does not literally fill in the missing scores, it implicitly does so via constraints imposed by the multivariate normality assumption.

Multiple Imputation

Multiple imputation is a second MAR-based approach that has become increasingly common in the literature. Unlike maximum likelihood, which estimates the model parameters directly from the available data, multiple imputation fills in the missing values before analysis. More specifically, a multiple imputation analysis consists of three phases: an imputation phase, an analysis phase, and a pooling phase. The imputation phase generates several copies of the data set (20 or more is a good rule of thumb; Graham, Olchowski & Gilreath, 2007), each of which contains a unique set of plausible replacement scores. In the analysis phase, the researcher performs the desired analysis on each complete data set. Finally, the pooling phase aggregates the parameter estimates and standard errors into a single set of results. Although the process of analyzing several data sets and pooling the results sounds tedious, many software packages fully automate this process. In this section, I provide a brief description of each phase, and additional

details are available elsewhere in the methodological literature (Enders, 2010; Schafer, 1997; Schafer & Graham, 2002).

The imputation phase creates a collection of complete data sets. Methodologists have developed algorithms for a variety of data structures and variable distributions (e.g., normally distributed variables, categorical variables, mixtures of continuous and categorical variables, multilevel data structures). I describe Schafer's (1997) data augmentation algorithm for normally distributed variables, but other algorithms follow very similar logic. Data augmentation is an iterative algorithm that repeatedly cycles between an imputation step (I-step) and a posterior step (P-step). The I-step sorts individuals into groups that share a common missing data pattern, then it uses regression equations to predict the incomplete variables from the complete variables. Because this process generates predicted scores that fall directly on a regression line or regression surface, the algorithm restores variability to the data by adding a normally distributed residual term to each score. The sum of a predicted score and a residual term replaces each missing data point.

Generating multiple sets of imputed values requires unique regression equations for each filled-in data set. The P-step uses Bayesian analysis techniques to generate alternate estimates of the regression model parameters. Conceptually, the P-step uses the filled-in data from the preceding I-step to define a sampling distribution for the variable means and the covariance matrix (the computational building blocks of the I-step regression equations).[2] The algorithm then uses Monte Carlo computer simulation to "draw" a new mean vector and covariance matrix from their respective distributions. The mean vector and covariance matrix carry forward to the next I-step where they serve as the basis for constructing new regression equations and new imputed values.

After generating imputations, the researcher analyzes each complete data set. For example, I later illustrate a multiple imputation analysis where I fit the growth model in Equation 2 to 50 imputed data sets. The analysis phase yields multiple sets of parameter estimates and standard errors, and the pooling phase subsequently uses Rubin's (1987) combining rules to aggregate these quantities into a single set of results. For any given parameter, the arithmetic average of the estimates serves as the multiple imputation point estimate. Pooling the standard errors is a bit more complex because the process incorporates two sources of sampling variation. The so-called within-imputation variance is the arithmetic average of the squared standard errors

$$\mathrm{W} = \frac{1}{m}\sum_{t=1}^{m} SE_t^2 \tag{6}$$

where m is the number of imputed data sets, and SE_t^2 is the squared standard error (i.e., sampling variance) from data set t. Because Equation 6 averages standard errors from the filled-in data, the within-imputation variance estimates the sampling error that would have resulted, had the data been complete.

The square root of Equation 6 would underestimate the standard error because it fails to incorporate the influence of the missing values. The between-imputation variance is effectively a correction factor that adds noise to account for this additional source of error. The between-imputation variance quantifies the variability of the parameter estimates across the m data sets, as follows

$$\mathrm{B} = \frac{1}{m-1}\sum_{t=1}^{m}(\hat{\theta}_t - \bar{\theta})^2 \tag{7}$$

where θ_t is the parameter estimate from data set t, and $\bar{\theta}$ is the average of the m estimates. Notice that Equation 7 is the usual formula for the sample variance, where parameter estimates replace score values. Equation 7 reflects missing data sampling error because differences among the imputed values across data sets solely determine its value (i.e., repeatedly analyzing a complete data set would yield $\mathrm{B} = 0$).

Finally, the within- and between-imputation variance combine to form a standard error, as follows.

$$SE = \sqrt{\mathrm{W} + \mathrm{B} + \mathrm{B}/m} \tag{8}$$

Consistent with a complete-data analysis, the standard error serves as the denominator of a test statistic (e.g., a t or z test).

Although multiple imputation and maximum likelihood make the same assumptions and tend to produce comparable estimates, imputation is arguably more complex to implement. For example, data augmentation and comparable algorithms produce results that are correlated from one computational step to the next. This implies that imputed values from one I-step will have a strong correlation with imputed values from the next (or the preceding) I-step. Because the goal of multiple imputation is to generate independent samples from a distribution of plausible replacement values, analyzing data sets from consecutive I-steps is inappropriate (doing so attenuates standard errors because the between-imputation variance is too small). Rather, the typical strategy is to generate a long sequence of computational cycles and save data sets at regular intervals. For example, I later illustrate an analysis where I generate 50 imputations by saving a data set at every 500th I-step. Determining the appropriate interval (i.e., the between-imputation iterations or thinning interval) is not necessarily straightforward and depends on the algorithm's convergence speed. Unfortunately, it is impossible to provide good rules of thumb because a number of data-specific characteristics influence convergence speed (e.g., sample size, number of variables, missing data rates, correlations among the variables). A number of authors provide illustrations of these diagnostic techniques (Enders, 2010; Schafer, 1997; Schafer & Olsen, 1998).

Multiple imputation is also difficult because it requires careful planning. Ideally, a single collection of imputed data files can accommodate all of the subsequent statistical analyses. For this to happen, the imputation phase must incorporate all of the variables from the analysis phase as well as any higher-order effects (e.g., interaction terms) that might be of interest. In addition the imputation algorithm must preserve any special features of the data structure—this is particularly important in longitudinal studies. For example, the depression data set is characterized by a common assessment schedule for all participants, such that every individual provides monthly assessments (i.e., a common covariance matrix

[2] To avoid delving into the mathematical details of Bayesian analyses, I use the term sampling distribution to describe the Bayesian concept of a posterior distribution. Although a sampling distribution and posterior distribution are analogous and often share the same shape (they generally do in this context), it is important to note that they are distinct concepts.

applies to all participants). For logistical reasons, researchers often implement longitudinal designs with person-specific data collection schedules, such that the interval between assessments varies across individuals (i.e., each assessment schedule requires a unique covariance matrix). These two data collection strategies require different imputation algorithms; standard algorithms such as the one I described above are appropriate for the former design, whereas multilevel algorithms are necessary for the latter. More generally, multilevel data sets—longitudinal or cross-sectional—require special imputation algorithms. Mistler and Enders (2011) describe multilevel imputation and provide a custom SAS macro for this purpose.

NMAR Analysis Methods

Recall that a NMAR mechanism occurs when the probability of missing data on a variable is related to the would-be value of that variable (e.g., the propensity for a missing depression score at a particular wave depends on the would-be value of that score). Methodologists have proposed a number of NMAR analysis models for longitudinal data, all of which augment the basic analysis with a model that explains the probability of missing data. I provide a brief description of three "classic" approaches: the selection model, the shared parameter model, and the pattern mixture model. Enders (2010, 2011) provides a more detailed description of these modeling frameworks, and Muthén, Asparouhov, Hunter, and Leuchter (2011) describe a number of interesting and promising extensions. A number of technical resources are available as well (e.g., Albert & Follmann, 2009; Diggle & Kenward, 1994; Hedeker & Gibbons, 1997; Little, 2009; Molenberghs & Kenward, 2007; Verbeke, Molenberghs, & Kenward, 2000; Wu & Carroll, 1988).

The Selection Model

Heckman (1976, 1979) originally proposed the selection model for regression analyses with NMAR missingness on the outcome variable, and methodologists have since extended this work to longitudinal models. The basic idea behind longitudinal selection modeling is to augment the growth model with additional regression equations that predict a set of binary missing data indicators (e.g., $R = 0$ if the outcome variable is observed, $R = 1$ if the outcome is missing). For example, the Diggle and Kenward (1994) selection model uses the repeated measures variables to predict the probability of missing data at a particular wave. To illustrate, Figure 2 shows a path diagram of the selection model that I later fit to the depression data. The diagram is largely the same as that in Figure 1 but incorporates two binary missing data indicators (the rectangles labeled R_2 and R_3). The model accommodates an MAR mechanism by incorporating lagged associations between the depression scores and the indicators (i.e., the regression of R_2 on Y_1 and the regression of R_3 on Y_2), and it incorporates an NMAR mechanism via the concurrent associations between the outcomes and the indicators (i.e., the regression of R_2 on Y_2 and the regression of R_3 on Y_3). Note that I use dashed arrows to differentiate the logistic regression equations from linear regressions.

Although it is not immediately obvious, the selection model requires strict and untestable assumptions that go beyond the missing data mechanism. For example, the logistic regressions that link the outcome variable at wave t to the corresponding missing data indicator are typically inestimable because the outcome variable is always missing whenever R equals one. These associations are only estimable by invoking strict distributional assumptions for the repeated measures variables, typically multivariate normality. Because these assumptions are so integral to model identification and estimation, even the relatively modest departures from normality that are common in the behavioral sciences can introduce substantial bias. The accuracy of the model also depends on correct specification of the dropout process. For example, notice that the model in Figure 2 uses only main effects to predict the missing data indicators, whereas interactive effects were actually responsible for dropout (i.e., low scoring individuals in the treatment group were more likely to quit the study, and high scoring individuals in the control group were more likely to drop out). Unfortunately, a misspecification such as this would likely introduce bias, although there would be no way of knowing that the model is misspecified. This underscores my previous assertion that NMAR models are not necessarily a reliable panacea for potential MAR violations.

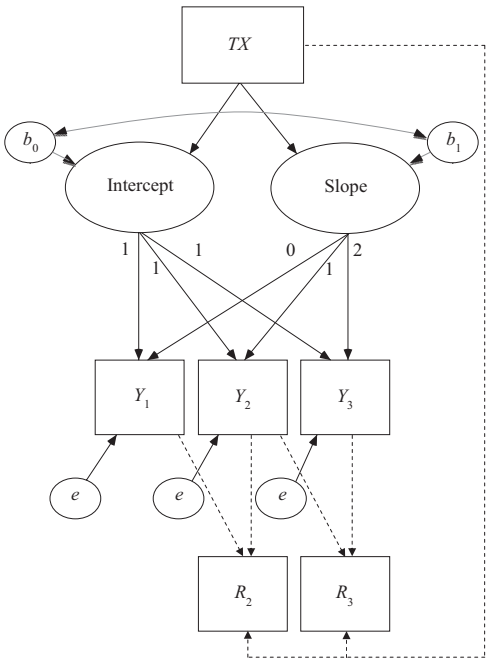

Figure 2. Path diagram of the longitudinal selection model. Solid arrows represent linear regressions, and dashed arrows denote logistic regressions.

The Shared Parameter Model

Like the selection model, the shared parameter model (Wu & Carroll, 1988) augments the growth curve analysis with logistic regression equations that predict a set of binary missing data indicators. However, the shared parameter model uses the individual growth curves (i.e., the b_{0_i} and b_{1_i} terms in Equation 2) as predictors of missingness. To illustrate, Figure 3 shows a path diagram of the model that I later apply to the depression data.

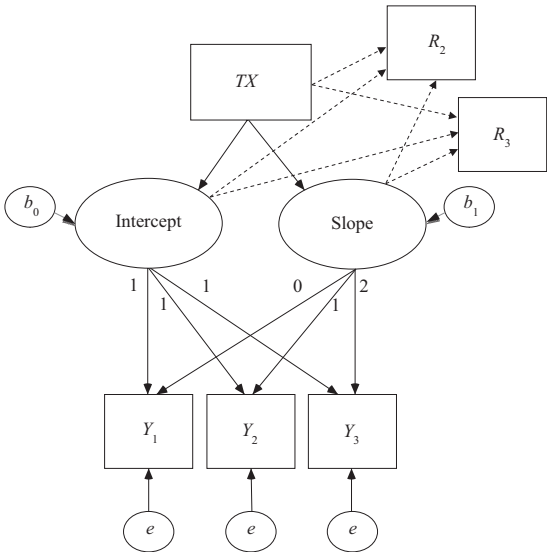

Figure 3. Path diagram of the longitudinal shared parameter model. Solid arrows represent linear regressions, and dashed arrows denote logistic regressions. To reduce visual clutter, the figure omits the latent variable covariance.

Notice that the growth factors rather than the repeated measures variables predict R_2 and R_3. Although the shared parameter model is conceptually similar to the selection model,[3] the logistic portion of the analysis provides a different explanation for attrition. For example, in the depression study, the shared parameter model posits that an individual's rate of change across the entire study (as opposed to a time-specific realization of depression) is predictive of dropout. Using the growth trajectories to predict the propensity for missing data simultaneously incorporates the entire set of repeated measures variables, including the would-be scores that are missing.

Like the selection model, the shared parameter model requires untestable assumptions that go beyond the missing data mechanism. Distributional assumptions are once again important, although this model requires multivariate normality of the individual intercepts and slopes. Additionally, the model assumes conditional independence between the repeated measures variables and the missing data indicators (i.e., controlling for the intercepts and slopes, the repeated measures variables do predict missingness). Finally, the shared parameter approach assumes that the logistic portion of the model accurately specifies the dropout mechanism. Again, violating one or more of these assumptions can produce biased parameter estimates.

The Pattern Mixture Model

Pattern mixture models are a third family of NMAR analyses. Procedurally, pattern mixture models are quite different from the previous approaches. The basic idea is to stratify the sample into subgroups that share a common missing data pattern and estimate the growth model separately within each pattern. Returning to Table 2, the depression data set has three missing data patterns: cases that drop out after the first wave, cases that drop out after the second wave, and cases that complete the study. The basic analysis would yield pattern-specific estimates of the model in Equation 2, and averaging across the missing data patterns would produce a population estimate of each parameter.

Note that there are different ways to specify a pattern mixture model and thus different ways to depict the model in a path diagram. For example, Hedeker and Gibbons (1997) and Muthén et al. (2011) use a set of dummy variables to define the missing data patterns, and they subsequently use these code variables to predict the growth factors. Applied to the depression example, the path diagram for this specification would be identical to that in Figure 1 but would incorporate additional predictors (the pattern dummy codes and the interactions between the code variables and the treatment indicator). The specification that I use in the later analysis examples is conceptually similar to a multiple group model, whereby the model in Figure 1 is fit to each missing data pattern. I favor the latter approach because it allows for group-specific model modifications.

A recurring theme with NMAR models is that they are only estimable after invoking assumptions about the unobserved score values. The same is true of the pattern mixture model, although its assumptions are quite different from those of the previous models. To illustrate, consider the subsample of individuals that quit the depression study after the baseline assessment (i.e., pattern 3 in Table 2). The growth curve model is underidentified with a single assessment, and most of the parameters are inestimable (e.g., the slope coefficients, the slope variance, etc.) Consequently, the pattern mixture approach requires researchers to specify values for the inestimable parameters. Because we usually lack the information to do so, borrowing estimates from another pattern is often the only viable strategy for estimating these models. For example, equating the inestimable slope parameters (i.e., β_1 and β_3) to the corresponding estimates from the complete cases is one option, and setting the slopes equal to the coefficients from the cases that drop out before the final wave is another option. The methodological literature describes several other alternatives (Demirtas & Schafer, 2003; Hedeker & Gibbons, 1997; Enders, 2010, 2011; Molenberghs, Michiels, Kenward, & Diggle, 1998; Thijs, Molenberghs, Michiels, & Curran, 2002; Verbeke et al., 2000). Perhaps not surprisingly, the accuracy of the user-supplied parameter values dictates the degree of bias in the pattern mixture model estimates.

Data Analysis Examples

Because there is no way to verify that an MAR or an NMAR mechanism holds for a particular analysis, methodologists often recommend that researchers should explore the stability of their substantive conclusions by fitting models with alternative assumptions to the same data (i.e., conduct a sensitivity analysis). In line with this recommendation, I performed a series of analyses that applied the MAR- and NMAR-based approaches to the depression data. Specifically, I used the Mplus 6 computer program (Muthén & Muthén, 1998-2010) to implement maximum likelihood estimation, multiple imputation, the selection model, the shared param-

[3] The structural similarities of the two models have prompted some authors to refer to the shared parameter model as a random coefficient selection model.

eter model, and the pattern mixture model. Although many packages now implement maximum likelihood and multiple imputation, Mplus is advantageous because it offers routines for implementing NMAR models. The data set is available for download at www.appliedmissingdata.com, and the appendixes contain the Mplus scripts for the analysis examples.[4]

The analysis examples estimate the growth curve model in Equation 2. In an intervention study, researchers are usually interested in assessing treatment group differences at the final assessment. Centering the temporal predictor relative to the final assessment (e.g., by fixing the values of the *TIME* variable to −2, −1, and 0) would address this aim because the main effect for the treatment group indicator (i.e., the β_2 coefficient) quantifies the group mean difference at the final wave. However, dealing with the inestimable parameters in the pattern mixture model is made easier by centering the *TIME* variable relative to baseline, such that its values equal 0 at the initial assessment, 1 at the 1-month follow-up, and 2 at the 2-month follow-up. This choice is somewhat arbitrary because algebraically manipulating the growth model parameters yields an estimate of the endpoint mean difference, as follows:

$$\begin{aligned}\mu_{Treatment} - \mu_{Control} &= [(\beta_0 + \beta_2) + 2(\beta_1 + \beta_3)] - [\beta_0 + 2\beta_1] \\ &= \beta_2 + 2\beta_3 \end{aligned} \quad (9)$$

where the first set of bracketed terms is the model-implied treatment group average at the final wave, the second set of bracketed terms is the corresponding control group average, and 2 is the value of the *TIME* variable at the final assessment. Mplus allows users to define new parameters that are functions of estimated parameters, and I used this feature to estimate Equation 9 and its standard error.

Complete-Data Analysis

As a starting point, I fit the growth model to the depression data before imposing missing values. Although a computer simulation study that draws repeated samples from a population is the correct way to assess the accuracy of a missing data handling routine, the complete-data estimates provide an approximate benchmark for evaluating the five approaches. The column labeled Complete in Table 3 gives the complete-data estimates for selected parameters. As seen in the table, the treatment group mean was slightly lower than the control group mean at baseline, but the difference was not significant, $\hat{\beta}_2 = -1.092$, $p = .353$. Control group depression scores decreased at a rate of approximately one fifth of a point per month, on average, but the slope coefficient was nonsignificant, $\hat{\beta}_1 = -.228$, $p = .684$. Most importantly, the slope difference (i.e., the group by time interaction effect) was significant, such that the treatment group scores decreased more rapidly than those of the control group, $\hat{\beta}_3 = -2.268$, $p = .002$. Substituting the appropriate estimates into Equation 9 produced a mean difference of −5.629 at the final assessment. To further illustrate these estimates, I used the regression coefficients to compute the model-implied growth trajectory for each treatment group. Figure 4 displays these simple slopes.

Maximum Likelihood Estimation

Turning to the incomplete data, I first used MAR-based maximum likelihood missing data handling to estimate the growth curve model. Appendix A gives the Mplus program for this analysis. Table 3 gives the estimates and the standard errors for selected parameters, and Figure 5 displays the model-implied growth trajectories. Consistent with the complete-data analysis, the maximum likelihood estimates indicated that the treatment group experienced greater reductions in depression than the control group. However, the analysis overestimated the control group growth rate ($\hat{\beta}_1 = -.854$ vs. $\beta_1 = -.228$) and underestimated the slope difference ($\hat{\beta}_3 = -1.667$ vs. $\beta_3 = -2.268$). Although the groups were significantly different at the final assessment, the biased slope parameters produced a mean difference that was approximately 21% smaller than the corresponding complete-data estimate.

Multiple Imputation

Next, I applied MAR-based multiple imputation to the depression data. Unlike maximum likelihood, which directly estimates the model parameters from the available data, multiple imputation fills in the missing values before analysis. Monitoring the imputation algorithm's convergence behavior is an important step in the imputation phase. A variety of graphical diagnostic procedures are available for this purpose (e.g., see Enders, 2010; Schafer, 1997; Schafer & Olsen, 1998), but I relied primarily on the potential scale reduction factor (Gelman, Carlin, Stern, & Rubin, 1995). The potential scale reduction factor from an initial diagnostic run suggested that the algorithm converged in fewer than 500 iterations. Consequently, I generated 50 imputations by saving a filled-in data set at every 500th imputation cycle. In practical terms, specifying 500 between-imputation iterations ensured that the imputed scores approximated independent random samples from a distribution of plausible replacement values. Appendix B gives the Mplus program for the imputation phase.

After generating the imputed data sets, I fit the growth model in Equation 2 to each data set and averaged the resulting estimates and standard errors using Rubin's (1987) pooling equations. Although this step sounds incredibly tedious, Mplus fully automates the process (see the program in Appendix C). In fact, the entire analysis and pooling phase took approximately one second on a laptop computer! Table 3 gives the estimates and the standard errors for selected parameters. A quick inspection of the table shows that multiple imputation and maximum likelihood produced nearly identical estimates (by extension, the multiple imputation simple slopes were comparable to those in Figure 5). Consistent with maximum likelihood, multiple imputation produced a significant mean difference that favored the treatment group, but this estimate was approximately 21% lower than the corresponding complete-data estimate.

The fact that maximum likelihood and multiple imputation produced comparable results was no surprise because the two approaches made identical assumptions (MAR and multivariate

[4] Note that the analyses are estimable with the free demonstration version of Mplus, which is available for download at www.statmodel.com.

Table 3
Growth Curve Parameter Estimates From the Data Analysis Examples

Parameter	Complete	Est.	*SE*	*p*
Maximum likelihood				
Control baseline mean (β_0)	35.876	36.125	0.843	<.001
Control slope (β_1)	−0.228	−0.854	0.710	.229
Treatment baseline difference (β_2)	−1.077	1.244	.387	
Treatment slope difference (β_3)	−2.268	−1.667	0.917	.069
Endpoint mean difference	−5.629	−4.411	1.804	.015
Multiple imputation				
Control baseline mean (β_0)	35.876	36.121	0.843	<.001
Control slope (β_1)	−0.228	−0.818	0.648	.207
Treatment baseline difference (β_2)	−1.092	−1.088	1.243	.381
Treatment slope difference (β_3)	−2.268	−1.678	0.878	.056
Endpoint mean difference	−5.629	−4.445	1.716	.010
Selection model				
Control baseline mean (β_0)	35.876	36.073	0.829	<.001
Control slope (β_1)	−0.228	−3.447	0.741	<.001
Treatment baseline difference (β_2)	−1.092	−1.037	1.246	.405
Treatment slope difference (β_3)	−2.268	−1.358	0.992	.171
Endpoint mean difference	−5.629	−3.752	1.971	.057
Shared parameter model				
Control baseline mean (β_0)	35.876	36.180	0.871	<.001
Control slope (β_1)	−0.228	−1.369	2.613	.601
Treatment baseline difference (β_2)	−1.092	−1.117	1.199	.352
Treatment slope difference (β_3)	−2.268	−1.715	0.921	.063
Endpoint mean difference	−5.629	−4.547	1.881	.016
Pattern mixture model				
Control baseline mean (β_0)	35.876	36.029	0.839	<.001
Control slope (β_1)	−0.228	−0.087	1.180	.941
Treatment baseline difference (β_2)	−1.092	−1.015	1.244	.415
Treatment slope difference (β_3)	−2.268	−3.964	1.393	.004
Endpoint mean difference	−5.629	−8.942	2.839	.002

normality). These methods tend to produce equivalent results, particularly when the imputation phase and the maximum-likelihood analysis use the same set of variables (Collins, Schafer, & Kam, 2001; Schafer, 2003). Consequently, there is usually no statistical basis for choosing between the two approaches, although maximum likelihood is typically easier to implement.

Selection Model

Next, I applied three NMAR analyses, beginning with the selection model. Recall that the selection model augments the growth curve analysis with a logistic regression model in which the repeated measures variables predict the probability of missing data at a particular wave. Figure 2 shows a path diagram of the model,

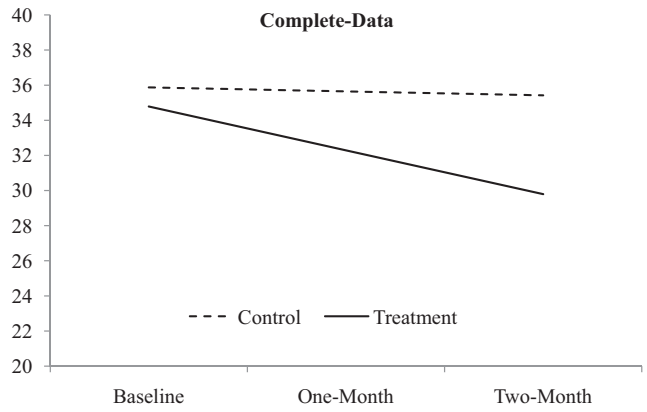

Figure 4. Model-implied growth trajectories (i.e., simple slopes) from the complete-data growth curve analysis.

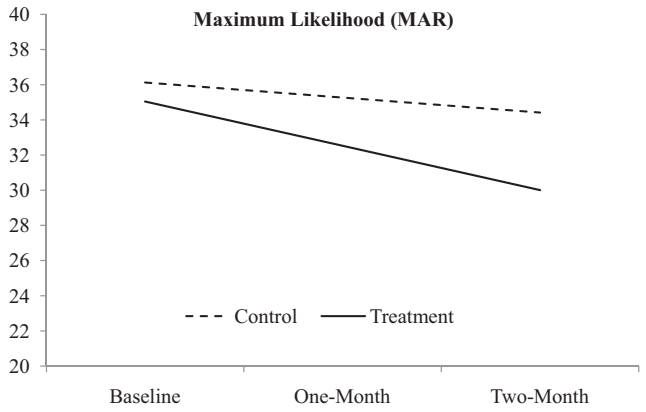

Figure 5. Model-implied growth trajectories (i.e., simple slopes) from the maximum likelihood (MAR-based) analysis.

and Appendix D gives the corresponding Mplus code. The coding of the missing data indicators warrants a brief discussion before proceeding. Diggle and Kenward (1994) originally proposed the selection model for longitudinal studies with permanent attrition (such is the case in this example). In this situation, the binary indicators are consistent with a discrete-time survival model where the codes take 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). Under this coding scheme, the logistic regression equations predict the probability of dropout at assessment t, given that a participant was in the study at the previous wave.

Although discrete-time codes are appropriate for the depression data, other coding schemes are appropriate for data sets that have a mixture of intermittent missingness and permanent attrition. Enders (2011) outlines three options: (a) implement discrete-time survival codes and assign intermittent missing values with zero code (i.e., complete), (b) treat each indicator as an independent Bernoulli trial, such that the code variables take on a value of zero at any assessment where the outcome is observed and take on a value of one at any assessment where the outcome is missing, and (c) treat each indicator as a multinomial logistic regression by adding a separate code for intermittent missingness. The first option effectively assumes that intermittent missingness is MAR, the second option assumes that the same underlying process causes intermittent missingness and permanent dropout, and the latter option treats intermittent missingness and permanent dropout as distinct processes. Enders (2011) described these coding schemes in more detail and illustrated their use on a real data set.

Implementing discrete-time survival indicators has bearing on model specification. Specifically, I imposed equality constraints on (a) the concurrent associations between the repeated measures variables and the indicators (i.e., the regression of R_2 on Y_2 was set equal to the regression of R_3 on Y_3), (b) the lagged associations between the depression scores and the indicators (i.e., the regression of R_2 on Y_1 was set equal to the regression of R_3 on Y_2), and (c) the associations between the treatment indicator and the missing data indicators (i.e., the regressions of R_2 and R_3 on the treatment variable were set equal). Readers who are interested in the rationale behind these constraints can consult Singer and Willett (2003) or other survival modeling resources.

Table 3 gives selected parameter estimates and standard errors from the selection model analysis, and Figure 6 shows the model-implied simple slopes. Unlike the MAR-based analyses, the selection model did not provide strong evidence for a treatment effect. As shown in the table, the model dramatically overestimated the control group growth rate ($\hat{\beta}_1 = -3.447$ vs. $\hat{\beta}_1 = -.228$) and underestimated the slope difference ($\hat{\beta}_3 = -1.358$ vs. $\hat{\beta}_3 = -2.268$). Consequently, the group mean difference was not significant at the final assessment ($p = .056$) and was roughly 33% smaller than the corresponding complete-data estimate.

Although not shown in the table, the logistic regression coefficients indicated that (a) treatment cases had a higher probability of dropout than control cases ($\hat{\beta} = -.853$, $SE = .409$, $p = .037$), (b) participants with higher depression scores at assessment t - 1 were more likely to leave the study at wave t ($\hat{\beta} = .138$, $SE = .035$, $p < .001$), and (c) cases with lower would-be scores at wave t were more likely to dropout at that wave ($\hat{\beta} = -.286$, $SE = .052$, $p < .001$). Recall that the logistic portion of the model is misspecified because it includes only main effects, whereas interactive effects were actually responsible for dropout (i.e., low scoring individuals in the treatment group were more likely to quit the study, and high scoring individuals in the control group were more likely to drop out). Model misspecification and normality violations (the repeated measures variables had skewness values between .50 and .70) were likely responsible for the rather large biases.

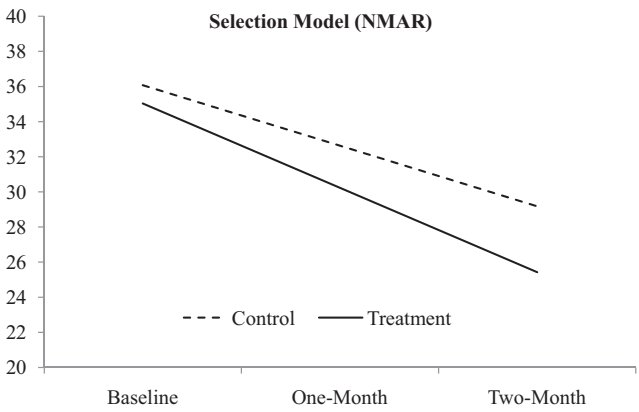

Figure 6. Model-implied growth trajectories (i.e., simple slopes) from the selection model (NMAR-based) analysis.

Shared Parameter Model

Recall that the shared parameter model relates the probability of missingness to the individual growth trajectories. To demonstrate this analysis, I fit the model from Figure 3 to the depression data. Consistent with the selection model, I used a discrete-time coding scheme for the missing data indicators (i.e., 0 = observed at wave t, 1 = dropout at wave t, missing = dropout at the previous wave). Further, I imposed equality constraints on (a) the associations between the treatment indicator and the missing data indicators, (b) the associations between the intercept latent variable and the missing data indicators, and (c) the associations between the slope latent variable and the missing data indicators. Appendix E gives the Mplus syntax for this analysis.

Table 3 gives selected parameter estimates and standard errors, and Figure 7 shows the model-implied simple slopes. As shown in the table and the figure, the shared parameter model overestimated the control group growth rate ($\hat{\beta}_1 = -1.369$ vs. $\hat{\beta}_1 = -.228$) and underestimated the slope difference ($\hat{\beta}_3 = -1.715$ vs. $\hat{\beta}_3 = -2.268$). Although the groups were significantly different at the final assessment, the biased slope parameters produced a mean difference that was approximately 19% smaller than the corresponding complete-data estimate. Recall that the MAR analyses produced comparable results. Interestingly, the partial regression coefficients from the logistic portion of the model were nonsignificant and suggested that attrition was unrelated to treatment group membership and to the individual intercepts and slopes. Like the selection model, the shared parameter model misspecified the dropout mechanism, and this misspecification likely contributed to the biases.

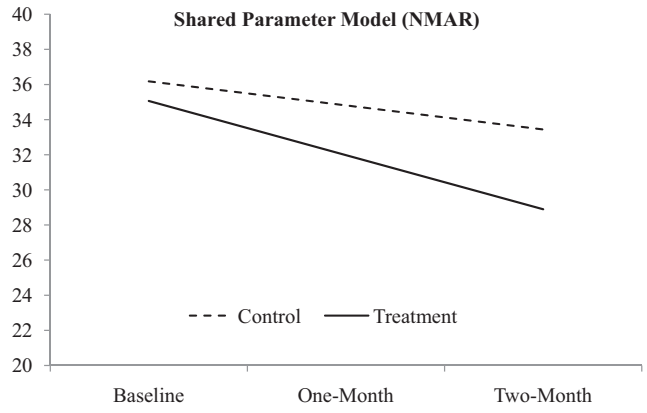

Figure 7. Model-implied growth trajectories (i.e., simple slopes) from the shared parameter model (NMAR-based) analysis.

Pattern Mixture Model

For the final example, I estimated a pattern mixture model. A pattern mixture model stratifies the sample into subgroups that share a common missing data pattern and estimates the growth model separately within each pattern. The depression data set has three missing data patterns: cases that drop out after the first wave, cases that drop out after the second wave, and cases that complete the study. The subsample of individuals who quit the study after the baseline assessment is problematic because several parameters are inestimable. To solve this problem, I combined all participants with missing data into a single group that I henceforth refer to as dropouts. This strategy implicitly assumes that all cases with missing data follow the same average trajectory, regardless of when they left the study. To identify the covariance structure, I further assumed that the complete cases and the dropouts shared the same variance estimates. These assumptions may or may not be tenable, but the model is otherwise inestimable. As an aside, researchers will often be faced with a large number of missing data patterns, some of which have only a few cases (e.g., a small group of participants with intermittent missing data at a single wave). As a practical matter, reducing the number of groups by aggregating patterns is often necessary (e.g., combining all patterns with intermittent missing values into a single group; grouping participants with intermittent values with the complete cases; combining patterns with similar observed means). Although there are no hard and fast rules for doing so, Enders (2011) illustrated the aggregation process on a real data set with nine sparse patterns.

It is important to note that the assumptions that I invoked for this analysis are just one possibility. For example, I could have separately estimated the model for all three groups, setting the inestimable slope parameters for pattern 3 equal to the estimates from the complete cases. Alternatively, I could have set the inestimable parameters from pattern 3 equal to the weighted average of the estimable parameters from patterns 1 and 2. In practice, applying a variety of identification strategies to the same data set is often a good idea, and a number of resources illustrate alternatives (Demirtas & Schafer, 2003; Enders, 2011; Hedeker & Gibbons, 1997; Muthén et al., 2011). As you will see below, I used the two-group model as a starting point and subsequently altered the would-be trajectory for the dropout group. Appendix F gives the Mplus syntax for the initial analysis.

The pattern mixture model yields group-specific estimates of the regression coefficients. Table 4 gives these estimates, and Figure 8 shows the corresponding simple slopes (note that the vertical axes are scaled differently than in previous graphs). I subsequently obtained population estimates by averaging across the missing data patterns, as follows:

$$\hat{\beta}_k = \hat{\pi}^{(1)}\hat{\beta}_k^{(1)} + \hat{\pi}^{(2)}\hat{\beta}_k^{(2)} \tag{10}$$

where $\hat{\pi}^{(1)}$ and $\hat{\pi}^{(2)}$ represent the proportion of completers and dropouts (.714 and .286, respectively). Table 3 gives the resulting parameter estimates and standard errors, and Figure 9 shows the corresponding growth trajectories. The pattern mixture model dramatically overestimated the slope difference ($\widehat{\beta} = -3.964$ vs. $\widehat{\beta}_3 = -2.268$), such that the treatment group appeared to improve more rapidly than it actually did. This bias produced a mean difference that was approximately 59% larger than the corresponding complete-data estimate.

The group-specific growth trajectories in Table 4 and Figure 8 provide insight into the performance of the pattern mixture model. Specifically, notice that the treatment cases in the dropout group showed a dramatic reduction in depression scores. Because the dropouts have only one or two observations, the large negative slope owes to the fact that the means decreased substantially between the first and the second assessment; the available-case means were $M_1 = 30.05$ and $M_2 = 23.55$, respectively. This poses a problem because the linear model effectively extrapolates the change between the first two waves to the change between the final two waves. Had the final assessment been observed, it seems unlikely that such a dramatic decrease would have persisted through the end of the study. The pattern mixture model is advantageous because it allows researchers to apply different assumptions about the would-be trajectory shapes. To accommodate the possibility of decelerating growth (i.e., rapid initial improvement that subsequently levels off), I estimated the model a second time after setting the *TIME* scores equal to 0, 1, and 1.1 in the dropout group. Although this choice was somewhat arbitrary, it modeled a situation where the change between the last two assessments was 10% of the change between the first two waves. Appendix G gives the Mplus syntax for the modified analysis.

Table 4
Group-Specific Estimates From the Pattern Mixture Model Analysis

Parameter	Est.	*SE*	*p*
Completers			
Control baseline mean (β_0)	35.260	0.791	<.001
Control slope (β_1)	−0.860	0.560	.684
Treatment baseline difference (β_2)	0.739	1.176	.353
Treatment slope difference (β_3)	−1.787	0.727	.002
Endpoint mean difference	−5.629	1.416	<.001
Dropouts			
Control baseline mean (β_0)	37.952	0.843	<.001
Control slope (β_1)	1.846	0.710	.229
Treatment baseline difference (β_2)	−7.900	1.244	.387
Treatment slope difference (β_3)	−9.405	0.917	.069
Endpoint mean difference	−4.411	1.804	.015

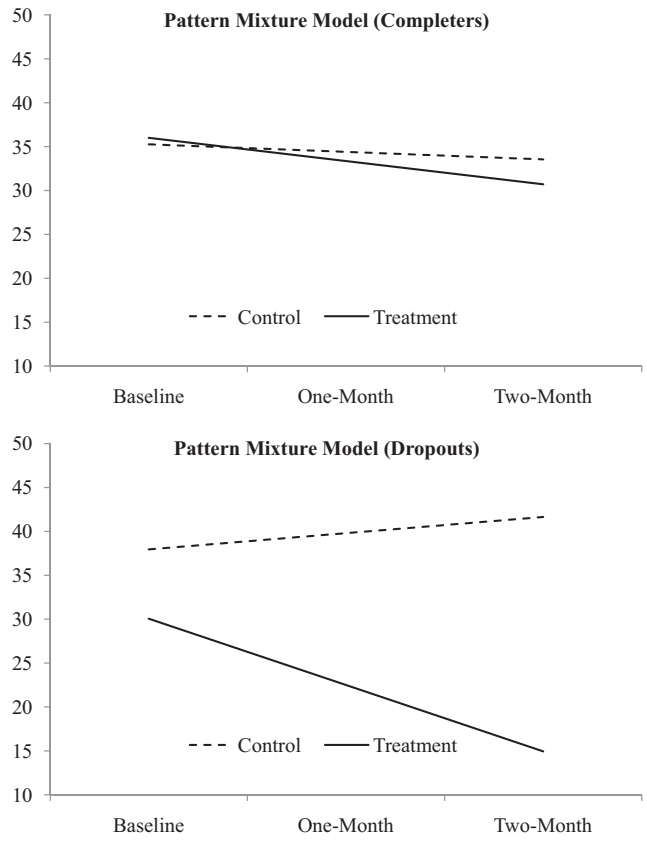

Figure 8. Pattern-specific growth trajectories (i.e., simple slopes) from the pattern mixture model analysis.

Because the interpretation of the growth model parameters differs between missing data patterns (e.g., for completers, β_1 represents the control group growth rate for the entire study, whereas the same parameter in the dropout group quantifies change between the first two assessments; Duncan, Duncan, & Stryker, 2006, pp. 31–35), pooling the group-specific estimates now makes little sense. However, averaging the group mean difference at the final wave is reasonable because this parameter has a common interpretation. The modified model with nonlinear growth in the dropout group produced a mean difference of −6.52, which is roughly 16% larger than the corresponding complete-data estimate.

Some readers may object to the fact that I arbitrarily altered the dropout group's parameter values. However, it is important to reiterate that the pattern mixture model requires researchers to invoke assumptions about the trajectory shapes that would have been observed, had the data been complete. Methodologists have argued that this facet of model specification is actually beneficial because it forces researchers to make their assumptions explicit. This is in contrast with the selection model and shared parameter model, both of which are only estimable via implicit distributional assumptions that are far from obvious. Because the dropout model extrapolates beyond the observed data, it is difficult to defend the initial estimates, particularly given that the treatment group slope approached the scale's minimum value at the final assessment. Consequently, model modification was absolutely necessary, although my set of *TIME* scores was just one option.

Analysis Summary and Recommendations

In the missing data literature, methodologists often recommend that researchers should conduct sensitivity analyses by fitting models with alternative assumptions to the same data. To illustrate this approach, I applied models with four sets of assumptions (MAR and three NMAR models) to the artificial depression data set. Because I created this data set specifically for illustration purposes, we have the luxury of knowing the "right answer" in the form of the complete-data estimates. However, this would not be the situation with real data. In practice, researchers must choose among a set of models that may or may not produce consistent estimates.

Although all of the analyses produced a mean difference that was in the expected direction, the magnitude of the effect varied across models; this is not an atypical result (Demirtas & Schafer, 2003; Enders, 2011; Foster & Fang, 2003). The most accurate model—the pattern mixture model—overestimated the true mean difference by 16%, whereas the other approaches underestimated the complete-data treatment effect, typically by about 20%. These discrepancies follow from the fact that the models made different *and incorrect* predictions about the unobserved data. The MAR analyses (maximum likelihood and multiple imputation) effectively assumed that the would-be depression scores at wave *t* could be inferred from treatment group membership or from the observed scores at previous assessments. Although this was true for many of the participants who dropped out after the baseline assessment, it was not true for cases who dropped out at the final assessment. In contrast, the NMAR models further assumed that an individual's propensity for missing data provided information about the would-be score values (or vice versa). Although this was true for the majority of participants, the models incorrectly specified the dropout mechanism. The analysis examples illustrate two important points. First, any missing data handling technique is only as good as the veracity of its assumptions. Second, NMAR-based analyses are not automatically superior to MAR-based anal-

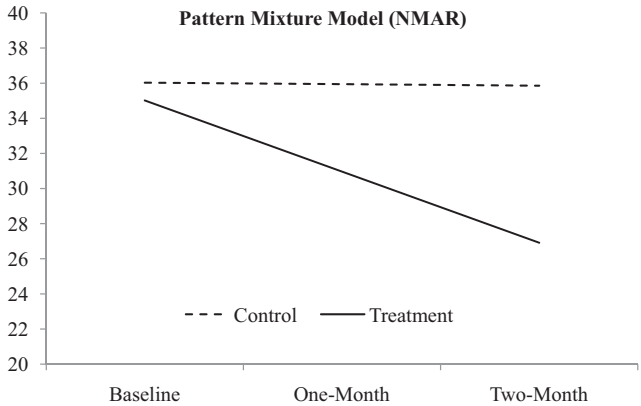

Figure 9. Model-implied growth trajectories (i.e., simple slopes) from the pattern mixture model (NMAR-based) analysis.

yses, even when attrition is largely consistent with an NMAR mechanism.

Unfortunately, there is no way to choose a single "correct" missing data handling technique because MAR and NMAR mechanisms invoke propositions that involve the unobserved scores. Consequently, researchers must select a model with the most defensible set of assumptions. I would argue that researchers should begin with an MAR analysis then move to one or more NMAR models. Although it is often a good idea to explore a variety of NMAR models, substantive considerations can guide model selection. For example, a selection model may be desirable when attrition is potentially related to the outcome variable at a single point in time (e.g., in a substance abuse intervention, participants who relapse may skip an assessment because they will screen positive). In contrast, the shared parameter model may be appropriate for situations where the developmental trajectories are probable determinants of missingness (e.g., in a quality of life study with cancer patients, individuals with rapidly decreasing trajectories are likely to dropout because they become too ill to participate). The shared parameter model is also useful when the outcome variable is either highly unreliable or highly variable over time (Albert & Follmann, 2009; Little, 1995). Finally, the pattern mixture model may be useful when the group-specific parameter estimates provide insight into one's substantive hypotheses (e.g., in an intervention study, it may be of interest to examine the response to treatment within each dropout pattern; see Muthén et al., 2011). It is important to reiterate that the observed data cannot inform model selection, so researchers generally need to provide logical arguments that defend their analytic choices.

Discussion

Missing data methodology has improved dramatically in recent years, and most popular computer programs now offer a variety of sophisticated options. Despite the widespread availability of theoretically justified analysis methods, researchers in many disciplines still rely heavily on subpar strategies that either eliminate incomplete cases or impute the missing scores with a single set of replacement values. Consequently, the purpose of this article was to provide readers with a nontechnical overview of some key issues from the missing data literature and to demonstrate several of the techniques that methodologists currently recommend.

Rubin's (1976) missing data mechanisms provide the basis for understanding when and why different missing data handling techniques work or fail. In the context of a longitudinal study, an individual's propensity for missing data at a particular wave may be related to predictor variables, to scores from previous assessments, or to the would-be scores from that assessment. Although less likely, missingness may be unrelated to other variables. The nature of the association between the probability of missing data and other variables (observed or unobserved) largely dictates the performance of a missing data technique. Older approaches (e.g., deleting incomplete cases) tend to make the restrictive assumption that missingness is unrelated to the data, and some methods yield bias regardless of the dropout mechanism (e.g., last observation carried forward). Given the shortcomings of these ad hoc approaches, I primarily focused on modern methods that assume MAR (i.e., missingness is related to observed variables) or NMAR (i.e., missingness is related to observed and unobserved variables).

MAR-based approaches (e.g., maximum likelihood and multiple imputation) are widely available in software packages, and these methods have the advantage of being relatively easy to implement. In addition, there are many situations where an MAR mechanism is quite reasonable. For example, consider a school-based study that examines depressive symptoms in a sample of adolescents. In this context, the primary source of attrition is often student mobility. Although mobility may be correlated with other variables such as socioeconomic status, it is difficult to argue that these correlates of missingness are strong enough to introduce bias, particularly after controlling for other variables in the analysis model. In other situations, it is reasonable to expect that MAR may be violated. For example, consider a longitudinal study of quality of life changes throughout the course of a clinical trial for a new cancer medication. In this scenario, it is likely that patients with rapidly decreasing quality of life scores are likely to leave the study because they die or become too ill to participate. NMAR-based analyses are a possible option because they allow for outcome-dependent missingness.

Unfortunately, there is no way to empirically assess the missing data mechanism because the MAR and NMAR assumptions both involve propositions about the unobserved (i.e., would-be) score values. In my experience, researchers in some disciplines are often quick to discount MAR-based analyses on grounds that outcome-dependent missingness is plausible. However, NMAR analyses are a far-from-perfect solution because they rely heavily on untestable assumptions that go beyond the missing data mechanism. For example, selection models and shared parameter models require strict distributional assumptions and proper specification of the dropout mechanism, and pattern mixture models require researchers to speculate about the trajectory shapes that would have been observed, had the data been complete. The fact that NMAR models require strict assumptions that limit their practical utility has led some methodologists to caution against their routine use (Demirtas & Schafer, 2003; Schafer, 2003).

In the end, there is no single "best" missing data handling technique that is guaranteed to work in all situations. Because MCAR-based analyses (e.g., deleting incomplete cases) are rarely if ever justified, researchers must choose between MAR and NMAR analysis methods. In the missing data literature, methodologists often suggest that researchers should conduct sensitivity analyses by fitting models with alternative assumptions to the same data. In line with this recommendation, it is often a good idea to begin with maximum likelihood estimation or multiple imputation and then examine whether the substantive conclusions are invariant across one or more NMAR models. In addition, researchers should provide logical arguments that defend their analytic choices, particularly given that their substantive conclusions are subject to one or more untestable assumptions.

References

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.

Bodner, T. E. (2006). Missing data: Prevalence and reporting practices. *Psychological Reports, 99,* 675–680.

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

Collins, L. M., Schafer, J. L., & Kam, C.-M. (2001). A comparison of inclusive and restrictive strategies in modern missing data procedures. *Psychological Methods, 6,* 330–351.

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.

Duncan, T. E., Duncan, S. C., & Strycker, L. A. (2006). *An introduction to latent variable growth curve modeling* (2nd ed.). Erlbaum: Mahwah, NJ.

Enders, C. K. (2001). The impact of nonnormality on full information maximum likelihood estimation for structural equation models with missing data. *Psychological Methods, 6,* 352–370.

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

Enders, C. K. (2011). Missing not at random models for latent growth curve analyses. *Psychological Methods, 16,* 1–16.

Enders, C. K., & Bandalos, D. L. (2001). The relative performance of full information maximum likelihood estimation for missing data in structural equation models. *Structural Equation Modeling: A Multidisciplinary Journal, 8,* 430–457.

Enders, C. K., & Gottschall, A. C. (2011). The impact of missing data on the ethical quality of a research study. In A. T. Panter & S. K. Sterba (Eds.), *The Ethics of Quantitative Methodology* (pp. 357–382). New York, NY: Routledge.

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.

Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (1995). *Bayesian data analysis.* Boca Raton, FL: Chapman & Hall.

Graham, J. W., Olchowski, A. E., & Gilreath, T. D. (2007). How many imputations are really needed? Some practical clarifications of multiple imputation theory. *Prevention Science, 8,* 206–213.

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.

Kwok, O.-M., Underhill, A. T., Berry, J. W., Luo, W., Elliot, T. R., & Yoon, M. (2008). Analyzing longitudinal data with multilevel models. An example with individuals living with lower extremity intra-articular fractures. *Rehabilitation Psychology, 53,* 370–386.

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.

Liu, G., & Gould, A. L. (2002). Comparison of alternative strategies for analysis of longitudinal trials. *Journal of Biopharmaceutical Statistics, 12,* 207–226.

Mistler, S. A., & Enders, C. K. (2011, August). *Applying multiple imputation to multilevel data sets: A practical guide.* Paper presented at the annual meeting of the American Psychological Association, Washington, DC.

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

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

Molenberghs, G., Thijs, H., Jansen, I., & Beunckens, C., Kenward, M. G., Mallinckrodt, C., & Carroll, R. J. (2004). Analyzing incomplete longitudinal clinical trial data. *Biostatistics, 5,* 445–464.

Muthén, B., Asparouhov, T., Hunter, A., & Leuchter, A. (2011). Growth modeling with non-ignorable dropout: Alternative analyses of the STAR*D antidepressant trial. *Psychological Methods, 16,* 17–33.

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.

Peugh, J. L., & Enders, C. K. (2004). Missing data in educational research: A review of reporting practices and suggestions for improvement. *Review of Educational Research, 74,* 525–556.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage.

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

Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys.* Hoboken, NJ: Wiley.

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

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

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

Schafer, J. L., & Olsen, M. K. (1998). Multiple imputation for multivariate missing-data problems: A data analyst's perspective. *Multivariate Behavioral Research, 33,* 545–571.

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, NY: Springer-Verlag.

Wilkinson, L., & Task Force on Statistical Inference. (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist, 54,* 594–604.

Wood, A. M., White, I. R., & Thompson, S. G. (2004). Are missing outcome data adequately handled? A review of published randomized controlled trials in major medical journals. *Clinical Trials, 1,* 368–376.

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.

(*Appendices follow*)

Appendix A

Mplus Maximum Likelihood Program

```
DATA:
! location of text data;
file = depression.dat;
VARIABLE:
! specify contents of data file;
names = tx dep1 dep2 dep3;
! select variables for analysis;
usevariables = tx dep1 - dep3;
! missing value code;
missing = all(-99);
ANALYSIS:
! mlr gives robust standard errors;
estimator = mlr;
MODEL:
! specify latent variable names and slope loadings;
icept Slope | dep1@0 dep2@1 dep3@2;
! treatment predicting baseline, parameter label in ();
icept on tx (b2);
! treatment predicting slope, parameter Label in ();
slope on tx (b3);
MODEL CONSTRAINT:
! create new parameter;
new(enddiff);
! use previous parameter labels to define new parameter;
enddiff = b2 + 2*b3;
```

Appendix B

Mplus Multiple Imputation Program (Imputation Phase)

```
DATA:
file = depression.dat;
VARIABLE:
! specify contents of data file;
names = tx dep1 dep2 dep3;
! select variables for analysis;
usevariables = tx dep1 - dep3;
! missing value code;
missing = all(-99);
ANALYSIS:
! specify saturated imputation model;
type = basic;
! random number seed;
bseed = 59765;
! convergence criterion for potential scale reduction factor;
bconvergence = .01;
OUTPUT:
```

(*Appendices continue*)

```
! print potential scale reduction factor values;
tech8;
DATA IMPUTATION:
! variables to be imputed;
impute = dep1 dep2 dep3;
! number of imputed data sets;
ndatasets = 50;
! file name for imputed data sets;
save = depimp*.dat;
! number of between-imputation iterations;
thin = 500;
```

Appendix C

Mplus Multiple Imputation Program (Analysis and Pooling Phase)

```
DATA:
! text file that contains the imputed data set names;
file = depimplist.dat;
! specify imputation data;
type = imputation;
VARIABLE:
! specify contents of data file;
names = tx dep1 dep2 dep3;
! select variables for analysis;
usevariables = tx dep1 - dep3;

ANALYSIS:

! mlr gives robust standard errors;
estimator = mlr;
MODEL:
! specify latent variable names and slope loadings;
icept slope | dep1@0 dep2@1 dep3@2;
! treatment predicting baseline, parameter label in ();
icept on tx (b2);
! treatment predicting slope, parameter label in ();
slope on tx (b3);
MODEL CONSTRAINT:
! create new parameter;
new(enddiff);
! use previous parameter labels to define new parameter;
enddiff = b2 + 2*b3;
```

(*Appendices continue*)

Appendix D

Mplus Selection Model Program

```
DATA:
! location of text data;
file = depression.dat;
VARIABLE:
! specify contents of data file;
names = tx dep1 dep2 dep3;
! select variables for analysis;
! r2 and r3 are indicator variables created by DATA MISSING;
usevariables = tx dep1 - dep3 r2 r3;
! missing value code;
missing = all(-99);
! define missing data indicators as categorical;
categorical = r2 r3;
DATA MISSING:
! variables used to create missing data indicators;
names = dep1 - dep3;
! define r2 and r3 as discrete time survival indicators;
type = sdropout;
binary = r2 r3;
ANALYSIS:
! mlr gives robust standard errors;
estimator = mlr;
! specify numeric integration for estimation;
algorithm = integration;
integration = montecarlo;
MODEL:
! specify latent variable names and slope loadings;
icept slope | dep1@0 dep2@1 dep3@2;
! treatment predicting baseline, parameter label in ();
icept on tx (b2);
! treatment predicting slope, parameter label in ();
slope on tx (b3);
! regress indicators on outcome at previous wave;
! values in () invoke equality constraints;
r2 on dep1 (1);
r3 on dep2 (1);
! regress indicators on outcome at current wave;
! values in () invoke equality constraints;
r2 on dep2 (2);
r3 on dep3 (2);
! regress indicators on treatment;
! values in () invoke equality constraints;
r2 on tx (3);
r3 on tx (3);
MODEL CONSTRAINT:
! create new parameter;
new(enddiff);
! use previous parameter labels to define new parameter;
enddiff = b2 + 2*b3;
```

(*Appendices continue*)

Appendix E

Mplus Shared Parameter Model Program

```
DATA:
! location of text data;
file = depression.dat;
VARIABLE:
! specify contents of data file;
names = tx dep1 dep2 dep3;
! select variables for analysis;
! r2 and r3 are indicator variables created by data missing;
usevariables = tx dep1 - dep3 r2 r3;
! missing value code;
missing = all(-99);
! define missing data indicators as categorical;
categorical = r2 r3;
DATA MISSING:
! variables used to create missing data indicators;
names = dep1 - dep3;
! define r2 and r3 as discrete time survival indicators;
type = sdropout;
binary = r2 r3;
ANALYSIS:
! mlr gives robust standard errors;
estimator = mlr;
! specify numeric integration for estimation;
algorithm = integration;
integration = montecarlo;
MODEL:
! specify latent variable names and slope loadings;
icept slope | dep1@0 dep2@1 dep3@2;
! treatment predicting baseline, parameter label in ();
icept on tx (b2);
! treatment predicting slope, parameter label in ();
slope on tx (b3);
! regress indicators on the intercepts;
! values in () invoke equality constraints;
r2 on icept (1);
r3 on icept (1);
! regress indicators on the slopes;
! values in () invoke equality constraints;
r2 on slope (2);
r3 on slope (2);
! regress indicators on treatment;
! values in () invoke equality constraints;
r2 on tx (3);
r3 on tx (3);
```

(*Appendices continue*)

```
MODEL CONSTRAINT:
! create new parameter;
new(enddiff);
! use previous parameter labels to define new parameter;
enddiff = b2 + 2*b3;
```

Appendix F

Mplus Pattern Mixture Model Program

```
DATA:
! location of text data;
file = depression.dat;
VARIABLE:
! specify contents of data file;
names = tx dep1 dep2 dep3;
! select variables for analysis;
! mdpatt is a new variable created by define;
usevariables = tx dep1 - dep3 mdpatt;
! latent categorical variable named pattern with two classes;
classes = pattern(2);
! latent class membership is defined by mdpatt variable;
knownclass = pattern(mdpatt = 1 mdpatt = 2);
DEFINE:
! create mdpatt variable that defines missing data patterns;
if (dep3 ne -99) then mdpatt = 1;
if (dep3 eq -99) then mdpatt = 2;
if (dep2 eq -99) then dep2 = _missing;
if (dep3 eq -99) then dep3 = _missing;
ANALYSIS:
! specify mixture model analysis;
type = mixture;
MODEL:
! global parameters for all classes;
%overall%
! specify latent variable names and slope loadings;
icept slope | dep1@0 dep2@1 dep3@2;
! treatment predicting intercepts and slopes;
icept slope on tx;
! intercept of latent categorical variable;
! label in () later used to compute class proportions;
[pattern#1] (p1logit);
! pattern 1 (completer) parameters with labels;
%pattern#1%
! latent variable means;
[icept] (b0p1);
[slope] (b1p1);
! treatment predicting intercepts and slopes;
icept on tx (b2p1);
slope on tx (b3p1);
! pattern 2 (dropout) parameters with labels;
%pattern#2%
! latent variable means;
```

(*Appendices continue*)

```
[icept] (b0p2);
[slope] (b1p2);
! treatment predicting intercepts and slopes;
icept on tx (b2p2);
slope on tx (b3p2);
MODEL CONSTRAINT:
! create new parameters;
new(pi1 pi2 b0 b1 b2 b3 enddiff);
! compute pattern proportions;
pi1 = exp(p1logit)/(exp(0) + exp(p1logit));
pi2 = exp(0)/(exp(0) + exp(p1logit));
! average estimates across patterns;
b0 = pi1*b0p1 + pi2*b0p2;
b1 = pi1*b1p1 + pi2*b1p2;
b2 = pi1*b2p1 + pi2*b2p2;
b3 = pi1*b3p1 + pi2*b3p2;
! endpoint mean difference;
enddiff = b2 + 2*b3;
```

Appendix G

Mplus Pattern Mixture Model Program (Modified)

```
DATA:
! location of text data;
file = depression.dat;
VARIABLE:
! specify contents of data file;
names = tx dep1 dep2 dep3;
! select variables for analysis;
! mdpatt is a new variable created by DEFINE;
usevariables = tx dep1−dep3 mdpatt;
! latent categorical variable named pattern with two classes;
classes = pattern(2);
! latent class membership is defined by mdpatt variable;
knownclass = pattern(mdpatt = 1 mdpatt = 2);
DEFINE:
! create mdpatt variable that defines missing data patterns;
if (dep3 ne −99) then mdpatt = 1;
if (dep3 eq −99) then mdpatt = 2;
if (dep2 eq −99) then dep2 = _missing;
if (dep3 eq −99) then dep3 = _missing;
ANALYSIS:
! specify mixture model analysis;
type = mixture;
MODEL:
! global parameters for all classes;
%overall%
! specify latent variable names and slope loadings;
icept slope | dep1@0 dep2@1 dep3@2;
! treatment predicting intercepts and slopes;
icept slope on tx;
```

(*Appendices continue*)

```
! intercept of latent categorical variable;
! label in () later used to compute class proportions;
[pattern#1] (p1logit);
! pattern 1 (completer) parameters with labels;
%pattern#1%
! latent variable means;
[icept] (b0p1);
[slope] (b1p1);
! treatment predicting intercepts and slopes;
icept on tx (b2p1);
slope on tx (b3p1);
! pattern 2 (dropout) parameters with labels;
%pattern#2%
! specify nonlinear growth with wave 3 loading;
icept slope | dep1@0 dep2@1 dep3@1.1;
! latent variable means;
[icept] (b0p2);
[slope] (b1p2);
! treatment predicting intercepts and slopes;
icept on tx (b2p2);
slope on tx (b3p2);
MODEL CONSTRAINT:
! create new parameters;
new(pi1 pi2 enddiff1 enddiff2 enddiff);
! compute pattern proportions;
pi1 = exp(p1logit)/(exp(0) + exp(p1logit));
pi2 = exp(0)/(exp(0) + exp(p1logit));
! endpoint mean difference for each pattern;
enddiff1 = b2p1 + 2*b3p1;
enddiff2 = b2p2 + 1.1*b3p2;
! average mean difference across patterns;
enddiff = pi1*enddiff1 + pi2*enddiff2;
```

Received July 17, 2011
Revision received August 18, 2011
Accepted August 19, 2011 ■