EVALUATING SHARED-PARAMETER MIXTURE MODELS FOR ANALYZING CHANGE IN THE PRESENCE OF NON-RANDOMLY MISSING DATA

Nisha C. Gottfredson

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for a Doctor of Philosophy degree in the Department of Psychology (Quantitative).

Chapel Hill
2011

Committee:

Daniel J. Bauer (Chair)

Patrick J. Curran

Andrea M. Hussong

Robert C. MacCallum

Antonio M. Morgan-Lopez

ABSTRACT

NISHA GOTTFREDSON: Evaluating Shared-Parameter Mixture Models for Analyzing Change in the Presence of Non-Randomly Missing Data
(Under the direction of Daniel J. Bauer)

Longitudinal researchers have been slow to adopt models for assessing the sensitivity of their results to potentially non-randomly missing data, opting instead to rely exclusively on more traditional approaches to modeling growth like latent curve modeling (LCM). Implicit in this choice is the strict assumption that missing data are missing at random (MAR). Failure to meet this assumption leads to inaccurate inferences regarding growth. A number of models for assessing the impact of non-randomly missing data on growth trajectory estimates have been presented over the past quarter century. These models are briefly discussed, and a new variation on some recently developed models is introduced. The shared parameter mixture model (SPMM) described here is preferable to some other models for a few reasons. Most notably, it approximates the dependence between the missing data process and the repeated measures without requiring an explicit specification of the missingness mechanism while simultaneously allowing conditional independence between the growth model and the missing data.

Performance of the SPMM is evaluated using simulation methodology across a range of plausible missingness mechanisms and across a range of longitudinal data conditions. SPMM performs well when the missing data mechanism is either latent

class- or growth coefficient- dependent. Fixed effect recovery is more robust than variance component recovery. The SPMM performs best with longer observation lengths and with erratically spaced missing data than with dropout.

Finally, this manuscript illustrates how the SPMM might be used in practiceby analyzing change over time in psychological symptoms of patients enrolled in psychotherapy.

Results are generally encouraging for SPMM performance under a range of simulated data conditions, and for feasibility with real data. Researchers who suspect the presence of random coefficient-dependent missing data are urged to consider using the SPMM to assess sensitivity of their model results to the MAR assumption.

ACKNOWLEDGEMENTS

This project was funded by NIDA (Grant F31 DA 026686). I would like to acknowledge Scott Baldwin and John Okiishi for generously providing the psychotherapy data that were analyzed in Chapter 4. I am indebted to all of the members of the L.L. Thurstone Lab, and particularly to the members of my dissertation committee, for invaluable critiques and suggestions as this work progressed. Most of all, I am grateful to my mentor, Dan Bauer.

DEDICATION

To my parents.

TABLE OF CONTENTS

LIST OF TABLES

Table

LIST OF FIGURES

Figure

CHAPTER 1

INTRODUCTION

Missing data is difficult to avoid in longitudinal social science research studies. Participant data can be missing for an entire wave of longitudinal data collection. Wave-level missingness might occur for relatively benign reasons (e.g., design-based missingness or inconvenient timing of data collection), or it might occur for reasons that are related to the research question of interest (e.g., death or hospitalization in a study of dementia-related outcomes; relapse in a study of alcohol treatment). Because longitudinal models are often used to make inferences about inter-individual variability in intra-individual change over time, longitudinal studies are vulnerable to bias resulting from a type of missing data that is uniquely troubling: missingness due to latent individual differences in growth trajectories (i.e., random coefficient-dependent missingness).

The studies presented in this manuscript evaluate a promising statistical technique for handling random coefficient-dependent missing data in longitudinal studies under a variety of real world data conditions. The introduction is organized as follows. First, the latent curve model (LCM) is described and different missing data mechanisms are defined within this context. Next, the relative strengths and weaknesses of popular models for handling non-randomly missing data will be discussed. The shared parameter mixture model (SPMM), a promising model for flexibly handling a variety of non-random missing data mechanisms, will then be described. Finally, hypotheses regarding the performance of the SPMM under several data conditions will be stated.

The Latent Curve Model and Missing Data

In the LCM, individual growth over time is modeled as follows (McArdle & Epstein, 1987; Meredith & Tisak, 1990; Bollen & Curran, 2006):

$$\begin{aligned} \mathbf{y}_i &= \boldsymbol{\Lambda}\boldsymbol{\eta}_i + \boldsymbol{\varepsilon}_i \\ \boldsymbol{\eta}_i &= \boldsymbol{\alpha} + \boldsymbol{\Gamma}\mathbf{x}_i + \boldsymbol{\zeta}_i \end{aligned} \qquad (1.1)$$

where $\mathbf{y}_i$ is a $T \times 1$ vector of repeated measures for individual i over T measurement occasions, $\boldsymbol{\eta}_i$ is a $m \times 1$ vector of latent growth scores (e.g., intercept, linear slope, quadratic slope) with a $m \times 1$ intercept vector $\boldsymbol{\alpha}$ and a $m \times 1$ vector of individual deviations $\boldsymbol{\zeta}_i$ that define the individual growth trajectories. A vector of q predictors/covariates $\mathbf{x}_i$ is related to individual growth factors through a $m \times q$ matrix of regression weights, $\boldsymbol{\Gamma}$. $\boldsymbol{\Lambda}$ is a $T \times m$ matrix of factor loadings (which are usually constrained by the analyst to define the shape of growth) that regress the repeated measures on latent growth factors, and $\boldsymbol{\varepsilon}_i$ is a $T \times 1$ vector of time-specific residuals. Growth factors are usually assumed to be conditionally multivariate normal $(\boldsymbol{\zeta}_i \sim N(\mathbf{0}, \boldsymbol{\Psi}))$, and they are generally allowed to covary with one another. Time-specific residuals are also assumed to be conditionally normally distributed ($\boldsymbol{\varepsilon}_i \sim N(\mathbf{0}, \boldsymbol{\Theta})$) and are often assumed to be independent (i.e., $\boldsymbol{\Theta}$ is a diagonal matrix); the last constraint may be relaxed. Further, it is assumed that the residuals are uncorrelated with the growth factors.

The equivalence of the latent curve model with mixed-effect, hierarchical, or multilevel growth models is well-established (Mehta & West, 2000; Bauer, 2003; Curran, 2003; Singer & Willett, 2003). The parameters within $\boldsymbol{\alpha}$ and $\boldsymbol{\Gamma}$ are the fixed effects, $\boldsymbol{\zeta}_i$ and $\boldsymbol{\varepsilon}_i$ are the random effects / residuals, and $\boldsymbol{\Psi}$ and $\boldsymbol{\Theta}$ contain the variance components and covariance parameters. Thus, issues discussed with respect to the latent curve model are equally applicable for growth models fit as mixed-effect, multilevel, or hierarchical linear models.

LCMs can be estimated using direct maximum likelihood (ML), resulting in unbiased growth parameter estimates when missing data are missing at random (MAR; Rubin, 1976) if the variables related to the missingness mechanism are measured and included in the data model (Arbuckle, 1996; Wothke, 2000; Enders, 2001). Similarly, if the causes of missingness are included in an imputation model prior to data analysis, then multiple imputation will lead to valid inferences under a MAR mechanism (Schafer, 2003; Rubin, 2004). However, when analyzing longitudinal data, there are many situations in which the MAR assumption for missing data would be untenable. As noted earlier, for instance, when studying change over time, it is possible that individual differences in growth are directly related to missingness probabilities (e.g., dropout in a longitudinal treatment study may be related to the progression of a disease; Demirtas & Schafer, 2003).

In general, the MAR assumption is violated when the cause of missingness is related to the outcome of interest and this cause is not included as a measured variable in the analytic model (or imputation model, if multiple imputation is used to account for a missing data process that is MAR). If the MAR assumption is violated, then the probability that a given repeated measure (y_{it}) is missing depends on the underlying value of y_{ti} itself, even after accounting for measured variables ($\mathbf{y}_i^o$ and $\mathbf{x}_i$, where $\mathbf{y}^o$ includes only the subset of observed repeated measures in $\mathbf{y}_i$). In this case, the missing data process is referred to as *outcome-dependent* and is 'nonignorable' (Rubin, 1976). Alternatively, the missing data can be said to be missing not at random (MNAR). If LCM is used to analyze data in which non-random missingness is present, bias may occur in fixed effect estimates ($\boldsymbol{\alpha}$ or $\boldsymbol{\Gamma}$) and variance estimates ($\boldsymbol{\Psi}$ or $\boldsymbol{\Theta}$).

Reflecting on Equation (1.1), there are two potential sources for non-ignorable outcome-dependent missingness in a longitudinal model: the random coefficients $(\boldsymbol{\eta}_i)$ that reflect inter-individual variability in change over time, or the time-specific residual errors $(\boldsymbol{\varepsilon}_t)$ that reflect intra-individual deviations from the individual's growth trajectory. Random coefficient-dependent missingness indicates a systematic trend of missingness across individuals (e.g., patients who experience little improvement in a clinical trial may drop out earlier than average), and error-dependent missingness indicates selection of observations within individuals (e.g., a participant in a daily diary study of pain may not report on particularly difficult days).

Modeling Growth in the Presence of Non-Randomly Missing Data

Any method for handling non-randomly missing data must somehow incorporate information about the missing data process into the model for the data. An in-depth review and illustration of several approaches for accomplishing this goal within longitudinal models was recently provided by Enders (2011). A more cursory review is offered here, leading up to a model that is particularly promising, the SPMM.

Selection models (Heckman, 1976; Heckman, 1979; Diggle & Kenward, 1984) and traditional shared parameter models (Wu & Carroll, 1988; Albert & Follman, 2009) require the analyst to specify an explicit model for the missing data and to condition the longitudinal model estimates on the missingness model. Whereas selection models condition the probability that a repeated measure is missing on the value of the repeated measure itself, shared parameter models condition the probability of missingness on individual growth trajectories. Thus, shared parameter models are particularly relevant for handling random coefficient dependent missingness. In shared parameter models, the growth parameters are 'shared' between the missingness model and the longitudinal model such that the missing data indicators are conditionally independent from the repeated measures after conditioning on the growth trajectories. Selection models and shared parameter models have the benefit of being conceptually straightforward, but they are heavily model-dependent and sensitive to misspecification of the missing data model (e.g., omitted covariates, misspecification of the form of missingness, or violations to distributional assumptions; Kenward, 1998; Winship & Mare, 1992; Vonesh, Greene, & Schlucher, 2006; Tsonaka, Verbeke, & Lesaffre, 2009).

Pattern mixture models (Little, 1993) and latent pattern mixture models (Roy, 2003) condition the longitudinal model parameters on observed or latent patterns of missingness so that a separate trajectory is estimated for each group of missing data patterns. In practice, this means estimating a growth model with an individual's missingness pattern included as a predictor (Hedeker & Gibbons, 1997; and possibly reducing the observed patterns to a smaller set of latent classes; Roy, 2003), similar to the multiple groups latent curve model for handling MAR missingness with an estimator that utilizes only sufficient statistics, a technique suggested by McArdle and Hamagami (1992). Group-specific trajectory estimates are aggregated to obtain a less biased trajectory estimate for the total population. The PMM and LPMM have the advantage that no *explicit* specification of the unknown missing data mechanism is required. However, these models suffer from the drawback that trajectories are (directly or indirectly) conditioned upon observed (and sample-dependent) missing data patterns. This may be problematic because the inclusion of missing data patterns as covariates in the trajectory model reduces the validity of ML-based inferences under an ignorable missing data mechanism (Demirtas & Schafer, 2003). According to Demirtas and Schafer, the inclusion of indicators of missingness (e.g., dropout occasion) as predictors in a growth model reduces the generalizability of inferences so that they are only valid under the specific growth mechanism that is implied by the model (including the precise form in which the indicators enter the model).[1]

A number of recent publications have combined the idea of a shared parameter model with the LPMM in order to induce conditional independence between the missing data indicators and the trajectory model of interest (Lin, McCulloch, and Rosenheck, 2004; Morgan-Lopez & Fals-Stewart, 2007; Beunckens, Molenberghs, Verbeke, & Mallinckrodt, 2008; Tsonaka et al., 2009; Muthén, Asparouhov, Hunter, & Leucter,

[1] Conditional independence models (i.e., shared parameter models) are preferable to the alternative pattern mixture approach because, according to Demirtas and Schafer (2003), conditioning the growth trajectory on missingness indicators limits the validity of the growth model beyond that implied by the (sample-dependent) model-implied mechanism. Pilot simulation work has verified that conditional independence models (e.g., shared parameter models) are more stable than conditional models like the PMM and LPMM.

2011). Lin et al. (2004) added a survival model for dropout. Others have taken Roy's (2003; LPMM) idea of using latent classes, initially viewed as a pattern reduction device, a step further by suggesting that the latent classes represent natural subgroups of individuals who differ qualitatively with respect to their missing data patterns and their growth trajectories. Some methodologists, however, have cautioned that seemingly distinct groups can often be estimated with such models even when heterogeneity is strictly continuous in nature, potentially resulting in misleading conclusions (Bauer & Curran, 2003; Sampson, Laub, & Eggleston, 2004; Bauer, 2007).

The Shared Parameter Mixture Model

The shared parameter mixture model draws on the models reviewed above to achieve several objectives. First, the model should not require the explicit specification of the missing data mechanism. The assumption underlying the first objective is that an analyst may have difficulty forming a correctly specified shared parameter model for the process underlying their missing data. Second, it is preferable to specify the growth model to be conditionally independent from the missing data indicators after accounting for exogenous variables and shared parameters (the idea behind traditional shared parameter models). Thus, an optimal model for random coefficient-dependent missingness would use a separate latent variable, distinct from the growth parameters, as the shared parameter between indicators for missingness and random growth coefficients.

To maximize flexibility in accounting for the missing data process without having to form an explicit model, the shared parameter should be discretely-distributed (i.e., a relatively small number of latent classes). The shared-parameter is a central part of the model because of its role in creating conditional independence between the trajectory and the missing data indicators (Tsonaka et al., 2009). Traditional shared parameter models rely on growth factors as the shared parameters, which are typically specified to be normally distributed. Misspecification of the shared-parameter distribution and its relation to other variables may lead to violation of the conditional independence assumption, leading to bias in trajectory estimates (Tsonaka et al., 2009). It is possible to circumvent this problem by conditioning the growth factors and the missing data patterns on discrete latent classes (the new shared-parameters) in order to approximate the unknown joint distribution between the growth factors and the missing data patterns. Indeed, latent mixture distributions are often used to semi-parametrically approximate unknown continuous densities (Heckman & Singer, 1984; Nagin, 1999). By using latent classes as an intermediary between growth factors and missing data patterns, it is possible to approximate the non-random missing data process semi-parametrically. The quality of the resulting approximation is the topic of Chapters 2 and 3 of this document.

Mathematically, the way that the SPMM factors the joint likelihood for the repeated measures and the missing data indicators can be expressed as follows:

$$f(\mathbf{y}_i, \mathbf{R}_i, \boldsymbol{\eta}_i, \mathbf{C}_i \mid \mathbf{x}_i) = \overbrace{f(\mathbf{y}_i \mid \boldsymbol{\eta}_i, \mathbf{x}_i)\, f(\boldsymbol{\eta}_i \mid \mathbf{C}_i, \mathbf{x}_i)}^{\text{Trajectory model, conditional on latent class, } C}\, \overbrace{f(\mathbf{R}_i \mid \mathbf{C}_i, \mathbf{x}_i)}^{\text{Semi-parametric missing data model}}\, f(\mathbf{C}_i) \qquad (1.2)$$

where $\mathbf{R}_i$ is a vector of missing data indicators (e.g., a T x 1 vector of binary indicators of missingness for every observation t where $r_{it} = 1$ if y_{it} is missing and $r_{it} = 0$ if y_{it} is observed). $\mathbf{R}_i$ could also be a one-number summary for the missingness patterns, as suggested by Roy (2007). When the number of repeated measures becomes large, estimation of SPMMs with binary indicators of missingness may become difficult. Examples of potential summary indicators are the number of total observations for individual *i* or the occasion of dropout for individual *i*.[2] Note that both the growth parameter and the missing data patterns are conditioned on the latent class variables, $\mathbf{C}_i$, as well as on the covariates $\mathbf{X}_i$. $\mathbf{C}_i$ is a set of latent, shared-parameter variables for the non-ignorable missing data mechanism. The effects of observed predictors may be included in the conditional distribution for $\mathbf{R}_i$ to account for a MAR mechanism, in order to make the model more efficient.

In the SPMM, covariates influence growth factors and missing data indicators directly, rather than indirectly via latent class probabilities. Although similar models presented in the literature allow covariates to affect class probabilities, this practice is not recommended for the SPMM because it complicates computation of the aggregate model parameters. Allowing covariates to predict class membership implies that marginal covariate effects depend on the values of the covariates themselves (Dantan, Proust-Lima, Letenneur, & Jacqmin-Gadda, 2008). Although true effects of covariates could be computed with some effort, estimation of the standard errors for covariate effects is intractable (Dantan et al., 2008).[3]

SPMMs can be specified as structural equation mixture models (SEMMs; Arminger, Stein, & Wittenberg, 1999; Dolan & van der Maas, 1998; Jedidi, Jagpal, & DeSarbo, 1997; Yung, 1997), and they can be estimated by ML with the EM algorithm

[2] Rose, von Davier, and Xu (2010) found empirical support for the practice of using summary indicators when implemented with a traditional PMM.

[3] The SPMM is meant for modeling continuously varying heterogeneity in a population, so classes are used as a statistical tool for approximating the missing data process. In a direct latent class pattern mixture modeling application in which discrete heterogeneity is assumed to exist, it may make sense to allow latent class predictors (e.g., Morgan-Lopez & Fals-Stewart, 2007).

using conventional software. With ML estimation, the optimal number of classes is determined by fitting a series of SEMMs, varying the number of latent classes present in each model, and comparing model fit. To estimate a SPMM, one specifies a mixture of latent curve models (i.e., growth mixture model (GMM); Verbeke & LeSaffre, 1996; Muthén & Shedden, 1999) with the form of growth that characterizes individual trajectories (e.g., linear, quadratic, piecewise), as shown below:

$$\begin{aligned} \mathbf{y}_i &= \mathbf{\Lambda}\boldsymbol{\eta}_i + \boldsymbol{\varepsilon}_i \\ \boldsymbol{\eta}_i &= \boldsymbol{\alpha}_k + \mathbf{\Gamma}\mathbf{x}_i + \boldsymbol{\zeta}_i \end{aligned} \tag{1.3}$$

where $\boldsymbol{\zeta}_i \sim N(\mathbf{0}, \mathbf{\Phi})$, $\boldsymbol{\varepsilon}_i \sim N(\mathbf{0}, \mathbf{\Theta})$, and the *k* subscript indicates a class-varying parameter. Unlike a conventional GMM, the SPMM jointly includes missing data indicators for the shared latent class variables via the equation

$$\mathbf{v}_i = \boldsymbol{\beta}_k + \mathbf{K}\mathbf{x}_i \tag{1.4}$$

where $\mathbf{v}_i$ is a vector of values for the linear predictor of $\mathbf{R}_i$, $\boldsymbol{\beta}_k$ is a vector of intercepts and $\mathbf{K}$ is a matrix containing the direct effects of the covariates $\mathbf{x}_i$ on the missingness indicators. For instance, if binary missing data indicators are present, then $\mathbf{v}_i$ might be specified as a vector of logits. An example path diagram for a SPMM with ten repeated measures and binary indicators of missingness is shown in Figure 1 (measurement parameters not annotated).

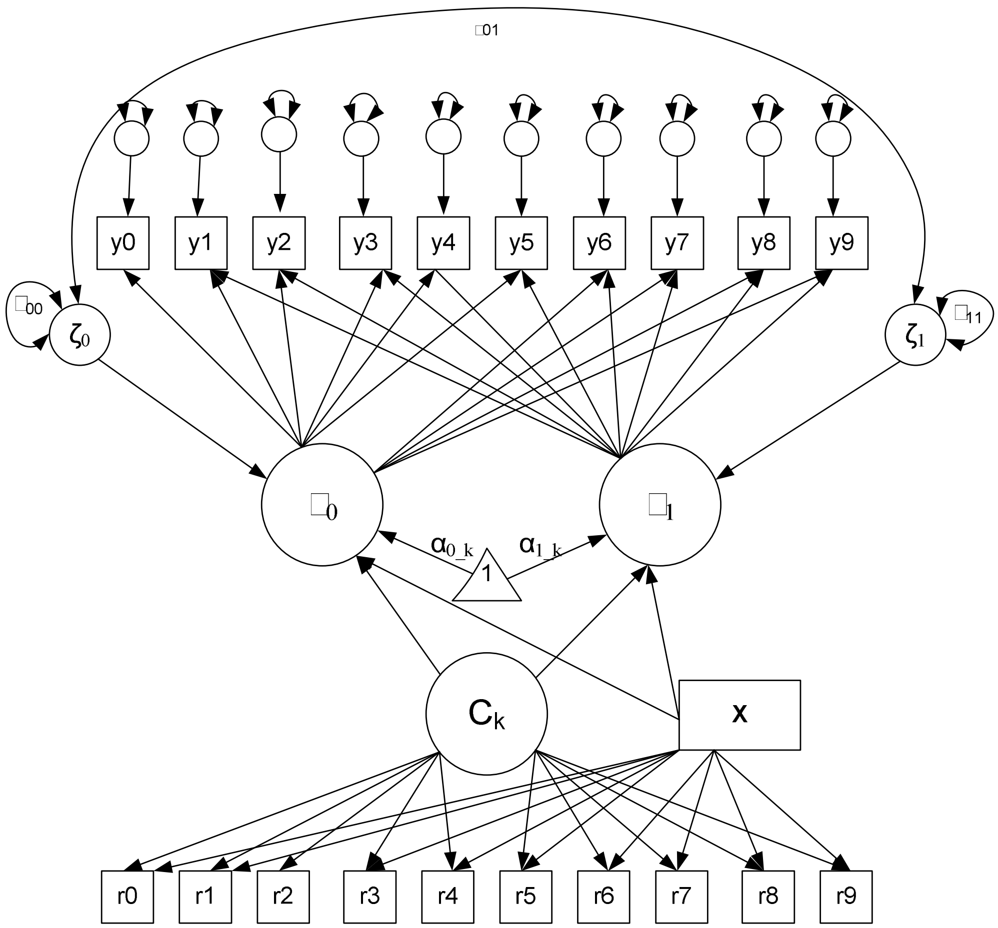


Figure 1. Path diagram of Binary SPMM. Circles represent latent variables and rectangles represent measured variables. Uni-directional arrows represent regression paths and bi-directional paths represent variances or covariances. The triangle represents the growth factor intercepts.

Note that the class-varying parameters in the SPMM of Equations (1.3) and (1.4) are $\boldsymbol{\alpha}_k$ and $\boldsymbol{\beta}_k$. Allowing these parameters to vary across classes enables the model to capture the dependence of the individual trajectories and the missing data. That is, joint differences in these parameter vectors allow certain types of trajectories (represented through $\boldsymbol{\alpha}_k$) to be associated with certain patterns of missing data (represented through $\boldsymbol{\beta}_k$). Although, in principle, some other parameters could also be permitted to vary across classes, limiting the number of class-varying parameters helps to retain parsimony, makes interpretation more straightforward (Dantan et al., 2008), and reduces the likelihood of some estimation problems (Hipp & Bauer, 2006).

When fitting a SPMM, one question is how many classes to include in the analysis. Numerous fit indices, including the Akaike Information Criterion (AIC, Akaike, 1974), the Bayesian Information Criterion (BIC; Schwarz, 1978), and many others, have been compared via simulation to determine the index with the optimal performance for GMMs (Lubke & Muthén, 2007; Tofighi & Enders, 2007). However, these studies have examined direct applications of mixtures and class recovery when true classes exist, whereas the goal of class enumeration is quite different here. The primary purpose of the latent classes in the SPMM is to explain the dependence between missing data patterns and growth parameters; the aim of class enumeration is to include enough latent classes to achieve this goal. The goal is not to determine the 'correct' number of latent classes. Therefore, it may be preferable to base class enumeration for SPMMs on the AIC, rather than the BIC, because AIC tends to prefer slightly more latent classes than BIC (McLachlan & Peel, 2000). That is, because the goal is to have a sufficient number of classes, more liberal selection indices, like AIC, may be preferable to more conservative ones, like BIC. On the other hand, including more latent classes than are necessary to achieve approximate conditional independence between the missing data indicators and the growth parameters may lead to imprecise estimates due to larger standard errors resulting from estimation of more parameters. From this perspective, the BIC may be the optimal index of fit.

Recovery of Fixed Effects and Variance Components

Once the number of classes have been selected, the next step is to aggregate over class estimates to obtain population level effects. Aggregate growth parameter means or intercepts are calculated by applying the following formula (Vermunt & van Dijk, 2001; Bauer, 2007):

$$\boldsymbol{\alpha} = \sum_{k=1}^{K} \pi_k \boldsymbol{\alpha}_k \tag{1.5}$$

where K is the total number of latent classes and π_k represents the class probability (mixing proportion, or weight) for class *k*. That is, class-specific means (for unconditional models) or intercepts (for conditional models), $\boldsymbol{\alpha}_k$, are weighted by their associated class probabilities, π_k , to obtain a population-average vector of growth factor means/intercepts.

Aggregate growth factor variances and covariances (or residual variances and covariances) can be calculated by combining the between-class covariance matrix (created by mean differences across classes) with the within-class covariance matrix, as shown below (Vermunt & van Dijk, 2001; Bauer, 2007):

$$\boldsymbol{\Psi} = \sum_{k=1}^{K} \sum_{j=k+1}^{K} \pi_k \pi_j (\boldsymbol{\alpha}_k - \boldsymbol{\alpha}_j)(\boldsymbol{\alpha}_k - \boldsymbol{\alpha}_j)' + \boldsymbol{\Phi} \tag{1.6}$$

For both Equations (1.5) and (1.6), aggregate estimates are obtained by substituting sample estimates for population parameters. Standard errors for the aggregate estimates can be computed via the delta method (e.g., Raykov & Marcoulides, 2004).

A number of researchers have noted that misspecification of the random-effects distribution can result in serious finite sample bias in variance component estimates, but that fixed effect estimates are typically unbiased (at least for normally distributed outcomes; e.g., Verbeke & Lesaffre, 1997; Litière, Alonso, & Molenberghs, 2007). Most research on the consequences of random effect misspecification has focused on assuming a continuous random effect distribution when the population generating model is a mixture; this manuscript deals with the opposite scenario. In a sense, the SPMM purposefully misspecifies the random effect distribution by imposing a mixture model when the true random effect distribution may be believed to be normal. Some recent manuscripts suggest that this type of misspecification may be just as troublesome for recovering variance component estimates (e.g., Sterba, 2010; Sterba, Baldasaro, & Bauer, 2010).[4]

In practice the 'true' population generating model is unknown, and a number of generating models are plausible. Thus, it is necessary to evaluate how well the SPMM is able to semi-parametrically approximate the population generating parameters under a variety of population generating mechanisms. This can be done via simulation methodology.

Evaluating the Performance of SPMMs

Enders (2011) showed that different approaches for accommodating MNAR data can provide widely varying substantive results. This is true in part because of the different assumptions required by each model, and in part because some models were created to handle slightly different forms of missingness (e.g., traditional selection

[4] The normal mixture that is implied by the SPMM is not exactly identical to the model analyzed in Sterba (2010) and Sterba et al. (2010). These studies used a discrete mixture, rather than a normal mixture, to approximate growth trajectories.

models were intended for outcome-dependent missingness and shared parameter and pattern mixture models were intended for random coefficient-dependent missingness). It should thus be emphasized that SPMM is intended to ameliorate parameter bias specifically due to random-coefficient-dependent missing data. Where SPMMs may fail is with outcome-dependent missing data (as opposed to strict random coefficient-dependent missing data). SPMMs cannot be expected to mitigate parameter bias associated with this type of problem entirely because, although the repeated measures are in part due to covariates and random coefficients, they also include residual error, a time-varying unmeasured effect. A similar observation may be made concerning more traditional pattern mixture models (with observed patterns) and it is noteworthy that these models have sometimes performed poorly with outcome-dependent missingness (Yang and Maxwell, 2009; Maxwell & Yang, 2010).

Overall, there has been very little published empirical work evaluating the performance of SPMM-type models. One exception is a study conducted by Morgan-Lopez and Fals-Stewart (2008), who simulated data under a discrete missing data mechanism (i.e., a mechanism in which there are a small number of groups with distinct missingness patterns). Specifically, they generated data from three groups: consistent attenders, dropouts, and erratic. The groups differed with respect to their probability of observation on each measurement occasion. In the population generating model, treatment predicted group membership, and both treatment group and latent group membership predicted the growth variables. The groups did not differ with respect to their overall growth trajectories, but they differed with respect to the treatment effect. The authors' analyzed the artificial data using: (a) the population-generating model (a latent class shared-parameter model with three latent classes), (b) a latent class shared-parameter model with two latent classes, and (c) a standard LCM. Morgan-Lopez and Fals-Stewart (2008) found that both the two- and three- class shared-parameter models resulted in acceptable parameter estimates, but that analyzing the data using a LCM resulted in unacceptable levels of bias when dropout rates were high and when class-specific treatment effects were well separated. The finding that even the two-class model recovered the aggregate parameters fairly well is consistent with the idea that one need not identify the true number of groups present in the data (in this case, three), so long as a sufficient number of groups is included in the model.

Study Overview

This study evaluated the performance of SPMMs for linear models of growth under a variety of population and modeling conditions. Predictors of growth are included to maximize the external validity of the results—in most real-world data analysis settings, researchers are more interested in making inferences about predictor effects than about modeling the shape of unconditional growth. Two versions of the SPMM are considered: the Binary SPMM and the Summary SPMM.

The major components of the study are described below.

<u>Study 1: Evaluating Performance of the SPMM under a Variety of Missing Data Mechanisms</u>

SPMMs are not designed to draw precise conclusions about the nature of the missing data mechanism; rather, they are designed to statistically approximate the joint distribution between observed missing data patterns and growth factors. As such, in order for the methodology to be useful in an applied context, it is necessary for the model to provide a good approximation to the missing data process under a wide range of missing data mechanisms in the population. Study 1 assessed the SPMM under a variety of missing data mechanisms: ignorable missingness (MAR), latent class-dependent non-ignorable missingness (SPMM-consistent missingness), growth coefficient-dependent missingness that is either monotonically (RC-MNAR-M) or non-monotonically (RC-MNAR-NM) related to the growth factor, and outcome-dependent missingness (OD-MNAR).

Two alternative SPMM specifications were evaluated: a model with a one-number summary of missingness (i.e., the number of repeated measures observed for each individual; a 'summary' SPMM) or a model with dichotomous missing data indicators for every repeated measure (a 'Binary' SPMM). Two models were considered for a couple of reasons. First, it is possible that relative model performance may differ by missing data mechanism. For instance, a summary SPMM might work well with a monotonic mechanism, but the same model might not work well with a non-monotonic (e.g., a U-shaped) missing data mechanism. The reason for this is as follows: if a mid-ranged random effect value is related to the lowest probability of missing data, with high probabilities of missingness on either tail of the random effect distribution, then the number of missing observations will be virtually uncorrelated with the growth factors. In this case, it might be more informative to use a Binary SPMM in order to adequately approximate the missingness process. The second reason to evaluate the performance of two models is practical: as the number of repeated measures increases, the computational feasibility of the Binary SPMM decreases.

Study 2: Evaluating the Effect of Missingness Mechanism Severity on SPMM Performance

Random coefficient-dependent missingness is expected to result in the most growth parameter bias when little information is available with which to accurately determine random coefficient estimates (conditions influencing growth factor determinacy are described below). Study 2 tested the hypothesis that the higher the determinacy of the random coefficients, the less severe the MNAR mechanism; as growth coefficients approach determinacy, they should become more like observed variables, and the MNAR mechanism should therefore approach a MAR mechanism. Longitudinal data characteristics that lead to good growth factor determinacy include: a high correlation between growth factors, a low proportion of unexplained variance (for both items and factors), and many repeated measures (e.g., Guttman, 1955; Mulaik & McDonald, 1978; Acito & Anderson, 1986).

When conducting an empirical research study, it is impossible to manipulate the correlation between growth factors, and control over the reliability of measures is limited. The number of repeated measures can be (relatively) easily manipulated by the researcher, so this was included as a measure of growth factor determinacy in Study 2. Five, 10, and 20 repeated measures were included as study conditions in order to simulate data across a realistic range of longitudinal studies. As recognized by Mulaik and McDonald (1978), infinitely adding observed variables does not additively contribute to factor determinacy. Thus, it might be expected that the difference between including five and ten repeated measures in a study is greater than the difference between ten and twenty indicators.

When data are missing, factor score determinacy will be influenced not only by the number of repeated measurement occasions, but also by the frequency with which the measures are actually observed. In Study 2, the proportion of missing data was manipulated so that the proportion is either .30 or .60. This range was chosen so that the proportion of MNAR missingness was high enough to substantially bias parameter estimates if MAR is assumed (Collins, Kam, & Schafer, 2001), and low enough to be realistic.

The nature of the random coefficient-dependent mechanism may influence factor score determinacy, holding constant the number of repeated measures observed. That is, even if the proportion of missing data were held constant, it would be reasonable to expect that an SPMM might have more trouble with a dropout mechanism than with an erratic missingness mechanism, particularly when the number of repeated measures is low. In instances of complete attrition, information about growth trajectories is concentrated in the initial part of the study, so growth factor estimates rely on a severely restricted range of information with which to provide inference about a parameter describing the entire time range in the study. The problem may be compounded if the proportion of unexplained variance is lowest at the intercept (i.e., if growth coefficients contribute more explained variance over time). However, the problem of dropout might be ameliorated if many repeated measures exist so that enough observations exist to develop a relatively precise trajectory estimate for individuals in the sample. Therefore, while holding the percentage of missing data constant, Study 2 varied the missingness mechanism to result in either erratic missingness or complete dropout. Simultaneously, the number of repeated measures were varied.

The same characteristics that influence growth factor determinacy in a LCM may also influence performance of SPMMs for a different reason. When a study collects a small number of repeated measures, when dropout is present (as opposed to erratically missing data), or when there is a high proportion of missing data, it is expected that fewer latent classes will be extracted from the data, thereby limiting the ability of the SPMM to fully account for dependence between the missingness mechanism and the growth factors.

Study 3: A Real-World Application of the SPMM

The final component of this manuscript is an empirical application of the SPMM to real data. Data are from a longitudinal study of patients while they were enrolled in psychotherapy. Past research has assumed that response to therapy treatment is independent from the dose of therapy received (Baldwin, Berkeljon, Atkins, Olsen, & Nielsen, 2009). Baldwin et al. showed that psychotherapy outcomes are not independent from dose (i.e., the number of psychotherapy sessions attended). They included total number of sessions attended as a predictor of psychotherapy outcomes in a growth model. An alternative strategy for modeling Baldwin et al.'s (2009) psychotherapy is to implement a SPMM, using number of sessions attended as a latent class indicator. The benefit of using the SPMM approach to modeling the psychotherapy data is that it provides a single model for the population that is not conditional on the number of sessions attended. Such information might be useful for understanding the average change over time that might be expected for an individual entering treatment, without a priori knowledge of the dose of psychotherapy that will be received. This knowledge would be useful in planning psychotherapy interventions.

The psychotherapy dataset analyzed in Study 3 is one example of many possible uses for SPMM methodology. In this case, participant data are structured using therapy sessions as the time metric. In a sense, patients who leave therapy before the final measurement occasion can be considered as dropping out of the study early; the full hypothetical trajectory (had the patients stayed in therapy until the end of the study) is unobserved for most participants. If time of dropout is related to participants' change trajectories, then the LCM-implied population average trajectories, and the variation around the average, will be biased toward the patients who stayed in therapy longer.[5]

[5] Trajectory estimates will be biased under an MAR approach assuming that the research question relates to the expected rate of change for individual patients enrolled in therapy, without a priori knowledge of their length in treatment.

Chapter 2

STUDY 1: QUALITY OF SPMM APPROXIMATION OF A VARIETY OF MISSING DATA MECHANISMS

Study 1 was designed primarily to assess SPMM performance under a variety of missing data mechanisms: MAR (i.e., ignorable) missingness, latent class-dependent missingness (i.e., SPMM-consistent missingness), growth coefficient-dependent missingness that is either monotonic (RC-MNAR-M) or non-monotonic (RC-MNAR-NM) with respect to the relationship between the growth coefficient and the probability of missingness, and outcome-dependent missingness (OD-MNAR).

Two alternative SPMM specifications were evaluated: a model with a one-number summary of missingness (i.e., the number of repeated measures observed for each individual; a 'Summary SPMM') and a model with binary missing data indicators for every repeated measure (a 'Binary SPMM'). The secondary purpose of Study 1 was to compare performance of the Binary SPMM with the Summary SPMM across a range of conditions under which the summary model would be expected to work as well as or better than the Binary model (MAR, SPMM-consistent, and RC-MNAR-M missingness mechanisms), and under conditions in which the binary model might provide some additional information regarding the missing data process with which to obtain less biased parameter estimates than the summary model could obtain (RC-MNAR-NM and OD-MNAR missingness mechanisms).

Hypotheses regarding model performance under the missingness mechanisms tested in Study 1 are listed below:

1. The LCM and the SPMM should both provide unbiased estimates of the average trajectories (i.e. $\boldsymbol{\alpha}$ and $\boldsymbol{\gamma}$) and variability around averages (i.e., $\boldsymbol{\Psi}$) when the missingness mechanism is MAR. The LCM should provide more efficient estimates than the SPMM since it relies on fewer parameters.

2. Both versions of the SPMM should provide less biased estimates of the population average trajectory than the LCM when the missing data mechanism is random coefficient dependent (i.e., SPMM-consistent, RC-MNAR-M, or RC-MNAR-NM). SPMM performance will be best when the non-ignorable mechanism is SPMM-consistent but it will provide a reasonable approximation to any random coefficient-dependent missingness mechanism.

3. SPMM-generated estimates will be less biased than LCM-generated estimates under an OD-MNAR mechanism to the extent that missingness depends on the random coefficients. However, given the presence of substantial residual variation, it is hypothesized that neither of the models will provide acceptable estimates under outcome dependent missingness.

Bias is considered for both fixed effects ($\boldsymbol{\alpha}$ and $\boldsymbol{\gamma}$) and variance components ($\boldsymbol{\Psi}$) in turn. With respect to the fixed effects, the greatest bias is anticipated for α_1 in the RC-MNAR conditions since the missing data generation mechanism is random slope dependent, and the greatest bias is anticipated for α_0 in the OD-MNAR condition since the missing data mechanism depends on levels of y_{ti}. For fixed effects, bias in standard error estimates is also examined.

Hypotheses regarding the relative performance of the Summary SPMM and Binary SPMM are:

4. The Binary SPMM may be better able to accommodate a non-monotonic mechanism than the Summary SPMM, but both models will be equivalently unbiased (or biased) in all other conditions.

5. The Summary SPMM should be more efficient than the Binary SPMM when both equivalently capture information about the missing data process (i.e., under a MAR mechanism, a RC-MNAR-M, and SPMM-consistent missingnes).

Data Generation

Parameter generating values were chosen to match the linear form of growth in a naturalistic psychotherapy study that was described in Baldwin et al. (2009).[6] In this study, patients were suspected to have left therapy, and thus stopped providing outcome information, as a function of their growth trajectories, suggesting a random-coefficient-process for study termination. Five hundred replicated samples of size 300 were generated for each missing data mechanism condition.[7]

For most of the conditions, data generation occurred in two steps. First, complete data $(\mathbf{y}_i^c)$ were generated, and then the observed repeated measures $\mathbf{y}_i^o$ were selected based on the missingness mechanism. An overall probability of 35% missingness was retained across all study conditions. Data on ten repeated measures were generated to be consistent with the following conditional LCM with a linear form:

[6] These data are analyzed in Chapter 4 using the SPMM.

[7] A sample size of 300 was chosen to balance between the complexity of the model (necessitating relatively large sample sizes) and the practical constraints of most psychological studies (necessitating relatively small sample sizes). Pilot research with larger sample sizes showed the same pattern of results, but with less variability and with more classes, on average.

$$
\begin{aligned}
y_{ti}^{c} &= \eta_{0i} + \lambda_t \eta_{1i} + \varepsilon_{ti} \\
\eta_{0i} &= \alpha_0 + \gamma_0 x_i + \zeta_{0i} \\
\eta_{1i} &= \alpha_1 + \gamma_1 x_i + \zeta_{1i}
\end{aligned}
\tag{2.1}
$$

where y_{ti}^{c} denotes complete data at time *t* for individual *i*, η_{0i} denotes the random intercept, λ_t is time ($\lambda_t = \{0,1,...,9\}$), η_{1i} is the random slope, and ε_{it} is the time-varying residual term, $\varepsilon_{ti} \sim N(0,180)$. The baseline intercept was set to $\alpha_0 = 69$ and random slope intercept was set to $\alpha_1 = -2.5$. Both were conditioned on the same binary time-invariant covariate, x_i $(x_i \sim Ber(.5))$, where the effect of the covariate is measured by regression parameters $\gamma_0 = 10$ (a moderate Cohen's *d* effect size of .52) and $\gamma_1 = -1.13$ (a moderate Cohen's *d* effect size of .42). Both growth factors were influenced by a randomly distributed disturbance term, ζ_{0i} and ζ_{1i}, respectively. The disturbances were distributed as follows:

$$
\begin{bmatrix} \zeta_{0i} \\ \zeta_{1i} \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{pmatrix} 375 & -10.38 \\ -10.38 & 7.18 \end{pmatrix} \right). \tag{2.2}
$$

Because time was coded to begin at zero, these values imply that the average individual with $x_i = 0$ begins the study with a score of 69 and declines by 2.5 units per time point, and the average individual with $x_i = 1$ begins the study with a score of 79 and declines by 3.63 units per time point. Individuals vary in their initial values and rates of change, as represented by the variation in random intercept and slope terms. Further, individuals whose intercepts are higher than average tend to decline at a faster rate than average, as implied by the negative covariance between ζ_{oi} and ζ_{1i} $(\rho = -.20)$.

Data for the SPMM-consistent, discrete missing data process was generated somewhat differently. Data in this condition were generated from three groups, each with a different probability of missingness (retaining an overall missingness probability of 35%). Each group also differed with respect to the average slope, but not with respect to the average intercept or covariate effects. Group 1 was characterized by a relatively flat rate of change ($\alpha_{11} = .7$ when $x_i = 0$) and a high probability of missingness ($p_1 = .68$).
Group 2 was characterized by a moderately negative rate of change ($\alpha_{12} = -2.5$ when $x_i = 0$) and a moderate probability of missingness ($p_2 = .35$). Group 3 was characterized by a steeply negative rate of change ($\alpha_{13} = -5.7$ when $x_i = 0$) and a low probability of missingness ($p_3 = .02$). Each group comprised 1/3 of the population. The overall population mean trajectory for this condition matched other conditions. Also, the population-level observed rate of change was -3.58, which is equivalent to the observed rate of change in the RC-MNAR-M condition . The within-class covariance matrix was specified as

$$\mathbf{\Phi} = \begin{pmatrix} 375 & -10.38 \\ -10.38 & 6.79 \end{pmatrix}. \tag{2.3}$$

Data deletion for the four SPMM-inconsistent conditions is described below.

MAR. Within each replication, the probability that a repeated measure was missing depended only on time (where t = 0 to 9).

$$p(r_{ti} = 1 \mid t) = \frac{e^{-.64+.8(t-4)}}{1+e^{-.64+.8(t-4)}} \tag{2.4}$$

Outcome dependent MNAR (OD-MNAR). The probability of missingness increased as the value of y_{ti}^{c} increased. A one standard deviation increase in y_{ti}^{c} was related to a 2.23 factor increase in the odds of item missingness. In Equation (2.5), σ_{yti} is the standard deviation of y_{ti}^{c}.

$$p(r_{it}=1 \mid y_{ti}) = \frac{e^{-.64+.8\left(\frac{y_{ti}-\bar{y}_{ti}}{\sigma_{y_{it}}}\right)}}{1+e^{-.64+.8\left(\frac{y_{ti}-\bar{y}_{ti}}{\sigma_{y_{ti}}}\right)}} \qquad (2.5)$$

Random coefficient dependent MNAR - monotonic process (RC-MNAR-M). The model for inducing monotonically increasing random coefficient-dependent MNAR was the following:

$$p(r_{it}=1 \mid \eta_{1i}) = \frac{e^{-.64+2\left(\frac{\zeta_{1i}}{\sqrt{\psi_{11}}}\right)}}{1+e^{-.64+2\left(\frac{\zeta_{1i}}{\sqrt{\psi_{11}}}\right)}}. \qquad (2.6)$$

For this condition, an SD increase in ζ_{1i} was related to a 7.39 factor increase in the odds of item missingness.[8]

Random coefficient-dependent MNAR - nonmonotonic process (RC-MNAR-NM). Missingness for the nonmonotonic random coefficient-dependence condition differed from the other non-random conditions because, although the mechanism was severe in that a strong relationship existed between the random coefficient and the probability of missingness, this relationship selected out information on both tails of the random effect distribution. Like the other conditions, overall missingness was fixed at 35%. A piece-

[8] MNAR severity is intentionally high; a more severe missing data mechanism provides a stronger test of SPMM (and LCM) performance.

wise function with five nodes was used to model missingness in this condition. In Equation (2.7), $Z_{\zeta_{1i}}$ represents the standardized value of the individual slope from the average slope ($Z_{\xi_{1i}} = \frac{\zeta_{1i}}{\psi_{11}}$):

$$\begin{aligned}
&\text{if } Z_{\zeta_{1i}} < -2 \text{ then } p(r_{ti}=1|\zeta_{1i}) = .6\\
&\text{if } -2 < Z_{\zeta_{1i}} < -1.5 \text{ then } p(r_{ti}=1|\zeta_{1i}) = -.2 - .4 * Z_{\zeta_{1i}}\\
&\text{if } -1.5 < Z_{\zeta_{1i}} < 0 \text{ then } p(r_{ti}=1|\zeta_{1i}) = .178 - .145 * Z_{\zeta_{1i}}\\
&\text{if } 0 < Z_{\zeta_{1i}} < 1 \text{ then } p(r_{ti}=1|\zeta_{1i}) = .178 + .156 * Z_{\zeta_{1i}}\\
&\text{if } 1 < Z_{\zeta_{1i}} < 2 \text{then } p(r_{ti}=1|\zeta_{1i}) = -.134 + .467 * Z_{\zeta_{1i}}\\
&\text{if } Z_{\zeta_{1i}} > 2 \text{ then } p(r_{ti}=1|\zeta_{1i}) = .8
\end{aligned} \quad . \tag{2.7}$$

Figure 2 illustrates the relationship between an individual's random slope deviation (ζ_{1i}) and the probability that an item is missing for the two RC-MNAR conditions.

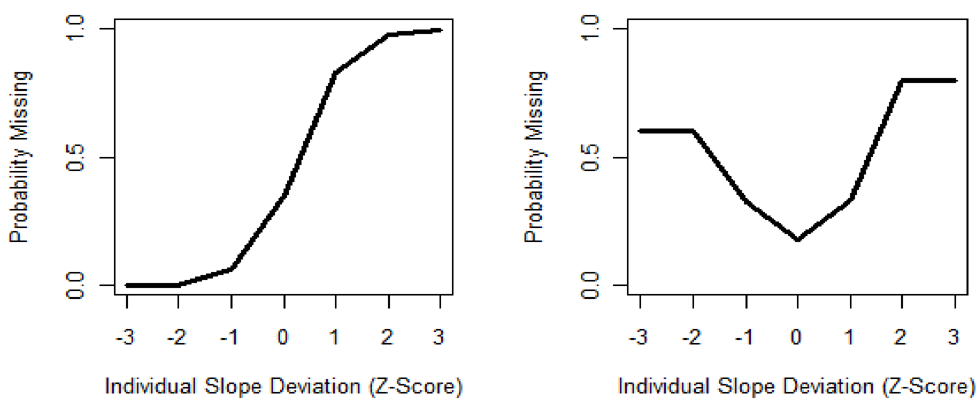


Figure 2. A Depiction of Monotonic and Non-Monotonic RC-MNAR Conditions: The Relations between ζ_{1i} and P($r_{ti}=1$)

Data Analysis

One- through five-class Binary and Summary SPMMs were estimated for each replicated dataset according to Equation (1.3). The missing data indicator in the Summary SPMM, the number of repeated measures observed for individual *i*, was treated as a continuous indicator and was assumed to be normally distributed within class. The assumption of normality is known to be violated, but with ten repeated measures the assumption violation is not egregious, and this assumption assists with computational feasibility (which is the impetus for using a summary indicator in the first place). Further, pilot research suggests that treating the summary indicator as a count variable (and modeling it with a Poisson distribution) does not improve model results.

For each replication, a class solution was removed if the solution was not positive definite, if the solution was a clear outlier upon visual inspection, or if the solution contained a class with probability less than .10.[9] Aggregate point estimates and delta-method standard error estimates were generated by Mplus (version 6) using Equations 1.4 and 1.5. Class enumeration was determined on a replication-by-replication basis; the models with the lowest AIC and BIC values were selected for comparison. A standard LCM, which assumes MAR, was also estimated for each replicated dataset for comparative purposes.

Table 1 reports rates of convergence to a positive definite solution and frequencies of positive definite solutions removed due to being an outlier or having a low class probability, by missing data mechanism and by model (Summary or Binary SPMM). The frequency with which one- through five-class solutions were selected by the AIC and BIC are also reported in Table 1.

[9] Solutions with small class proportions tend to produce very large standard error estimates that would in practice be rejected in favor of a solution with fewer classes, regardless of information criteria. Preliminary analyses indicated that solutions containing very small classes produced variance component estimates that were more upwardly biased than the estimates produced by solutions with more equal class proportions. Furthermore, Lubke and Neale (2006) found that small class sizes lead to difficulty in correct model detection, when a correct model exists.

As shown in Table 1, estimating up to five classes appears to have been more than sufficient for reaching conditional independence between growth factors and missing data indicators, at least as suggested by the AIC and BIC. Many high-class solutions were removed due to low class proportions, particularly when the missing data mechanism was MAR or OD-MNAR.

As expected, the BIC consistently chose fewer classes than the AIC. This is because the BIC contains a penalization for the number of independent observations (i.e., individuals, *n*) in the sample and the AIC does not:

$$\begin{aligned} AIC &= 2q - 2\ln(L) \\ BIC &= q\ln(n) - 2\ln(L) \end{aligned}, \qquad (2.8)$$

where *q* denotes the number of parameters and *L* is the likelihood.

Figure 3 shows the AIC- and BIC-based distributions of class enumeration for the MAR conditions in Study 1. Both the AIC and the BIC choose one class most of the time (59.00% for the AIC and 72.80% for the BIC) and both choose more than one class a good portion of the time, considering that it is known that a single class generated the sample data. This finding implies that neither the AIC nor the BIC should be used as an empirical test for whether MAR is a reasonable assumption for missing data, and it is a replication of earlier findings (e.g., Bauer & Curran, 2003; Tofighi & Enders, 2008).

Table 1. *Rates of Convergence to a Proper Solution, Solution Deletion, and AIC and BIC Model Selection for Study 1*

	Full SPMM						Summary SPMM					
Classes	Converged	Low π_k	Outlier	Remain	AIC	BIC	Converged	Low π_k	Outlier	Remain	AIC	BIC
MAR Mechanism												
1	468	NA	0	468	295	364	499	NA	0	499	257	305
2	467	42	0	425	79	55	498	121	10	367	88	74
3	466	155	2	309	35	20	492	236	4	252	78	78
4	459	321	0	138	46	25	476	342	0	134	51	38
5	457	409	0	48	13	4	469	445	0	24	7	4
SPMM Consistent Mechanism												
1	500	NA	0	500	0	0	500	NA	0	500	0	0
2	500	0	1	499	0	176	500	0	0	500	0	1
3	500	0	1	499	368	220	500	0	0	500	186	195
4	499	254	15	230	91	73	497	31	21	445	233	227
5	500	410	0	90	41	31	500	279	0	221	81	77
RC-MNAR-M Mechanism												
1	500	NA	0	500	0	0	500	NA	0	500	2	2
2	500	0	1	499	6	11	500	0	0	500	0	0
3	493	0	5	488	210	415	495	0	2	493	15	45
4	493	99	8	332	198	0	483	2	9	472	332	305
5	365	179	6	180	86	74	425	125	3	297	151	148
RC-MNAR-NM Mechanism												
1	500	NA	0	500	109	171	500	NA	0	500	98	112
2	500	140	0	360	195	304	500	136	0	364	196	251
3	498	257	1	240	118	6	499	259	0	240	173	117
4	479	367	0	112	50	0	461	407	0	54	25	13
5	447	403	0	44	28	19	387	103	0	10	8	7

	Full SPMM						Summary SPMM					
Classes	Converge	Low π_k	Outlier	Remain	AIC	BIC	Converge	Low π_k	Outlier	Remain	AIC	BIC
OD-MNAR Mechanism												
1	500	NA	0	500	1	2	500	NA	0	500	356	428
2	500	0	1	499	93	488	500	0	0	500	0	1
3	500	89	5	406	222	2	411	54	0	357	64	1
4	500	230	0	270	174	0	496	338	2	156	70	63
5	177	152	0	25	10	8	177	156	0	21	10	7

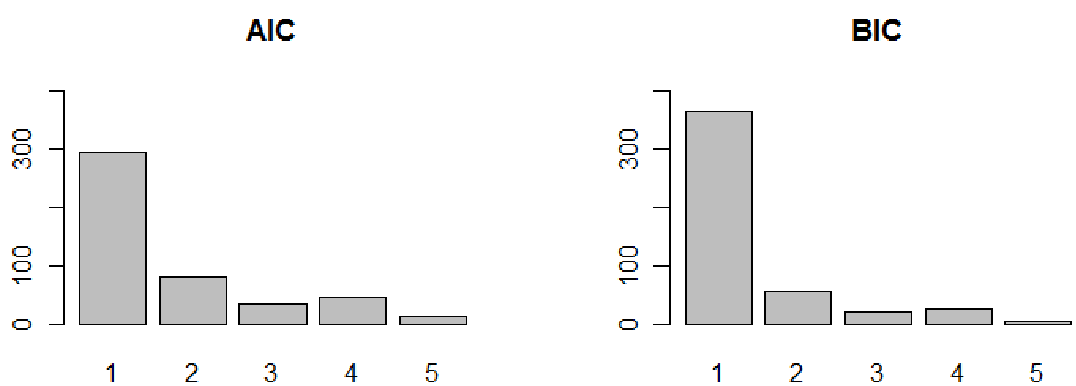


Figure 3. Class enumeration based on AIC (left) and BIC (right) when missing data are MAR

It is encouraging to note that a single class was *never* selected by the AIC or BIC with a Binary SPMM and a single class was very rarely selected with a Summary SPMM when the missing data mechanism was SPMM Consistent or RC-MNAR-M. The frequency distribution of class enumeration for the OD-MNAR condition is somewhat surprising; a single class was never selected by the AIC, and a single class was only chosen once by the BIC. It appears that the non-random censoring of the y_{ti} values led to a skewed distribution that resulted in multiple classes being selected. Also, OD-MNAR contains a RC-MNAR mechanism: individual differences in intercept, and to a lesser extent slopes, explain some variation in y_{ti}. Indeed, more classes were selected on average when the missing data mechanism was OD-MNAR than when the missing data mechanism was RC-MNAR-NM.

Standardized bias (SB) and root mean squared error (RMSE) were used as performance criteria for evaluating bias and precision of the fixed effect and variance component estimates from the LCM and SPMM in Study 1. Appendix A presents additional results comparing standardized, raw, and relative bias measures. SB was calculated as follows, where $\hat{\theta}_j$ is the estimate for θ in the *j*th repetition, and *N* is the total number of replications that are properly converged:

$$SB = \frac{100}{SE(\hat{\theta}_j)} * \left(\frac{\sum_{j=1}^{N} \hat{\theta}_j}{N} - \theta \right) \tag{2.9}$$

SB measures the magnitude of parameter bias as a percentage of the standard error for each parameter. It can be interpreted as the distance (in percentage of standard deviation units) that the estimate is off from the true parameter (Collins et al., 2001). According to Collins et al., SB values falling within ± 40% are considered 'acceptable.' This is equivalent to bias within ± .4 SD units for parameter estimates. Because SB is scaled by each parameter's standard error, it is useful in this study for comparing bias in parameter estimates across missing data conditions.

RMSE is a measure of the variation / imprecision of estimation that was calculated as follows:

$$\text{RMSE} = \sqrt{\frac{\sum_{j=1}^{N}(\hat{\theta}_j - \theta)^2}{N}}. \tag{2.10}$$

Accuracy of inferences related to predictor effects and growth factor means were further assessed by examining the ratio between the standard error estimates and the true, empirical standard deviations of the sampling distribution for each point estimate.

Study 1 results are presented below, organized by hypotheses. First, the LCM is compared with the Binary SPMM (according to the first set of hypotheses), and then the Binary SPMM is compared with the Summary SPMM (according to the second set of hypotheses). The LCM is compared with the Binary SPMM first because the Binary SPMM is similar to most latent mixture model formulations that have been presented in the literature for handling non-randomly missing data (e.g., Roy, 2003; Morgan-Lopez & Fals-Stewart, 2007, 2008). The Summary SPMM is compared with the Binary SPMM because it was introduced later in time in an effort to reduce computational complexity (e.g., Roy, 2007).

Results for Hypothesis 1: Trajectory Recovery under MAR

The first hypothesis posited that both LCM- and SPMM-implied trajectories would be equivalently unbiased in the fixed effects under a MAR mechanism, but that the LCM would be more precise than the SPMM.

Table 2 compares SB and RMSE of fixed effect trajectory estimates implied by the LCM, and by the (Binary) SPMM (both the best / lowest AIC and best / lowest BIC are reported), and Figure 4 shows that the average LCM- and SPMM-implied trajectories are both indistinguishable from the generating model.

Table 2. *Bias and Efficiency of Trajectory Recovery under a MAR Mechanism*

	LCM		SPMM (Best AIC)		SPMM (Best BIC)	
	SB (%)	RMSE	SB (%)	RMSE	SB (%)	RMSE
Fixed Effects						
Conditional Intercept (α_0)	4.89	1.84	.55	1.82	0	1.83
Conditional Slope (α_1)	0	.33	0	.35	2.86	.35
Intercept Predictor (γ_0)	-1.89	2.65	0	2.68	.75	2.71
Slope Predictor (γ_1)	2.22	.46	4.26	.47	2.13	.48
Variance Components						
Intercept Variance (ψ_{00})	-10.30	37.43	-16.49	56.01	-15.84	49.22
Slope Variance (ψ_{11})	-9.82	1.12	-17.76	1.69	-16.67	1.50
Covariance (ψ_{01})	2.28	4.82	10.73	6.68	8.98	6.09

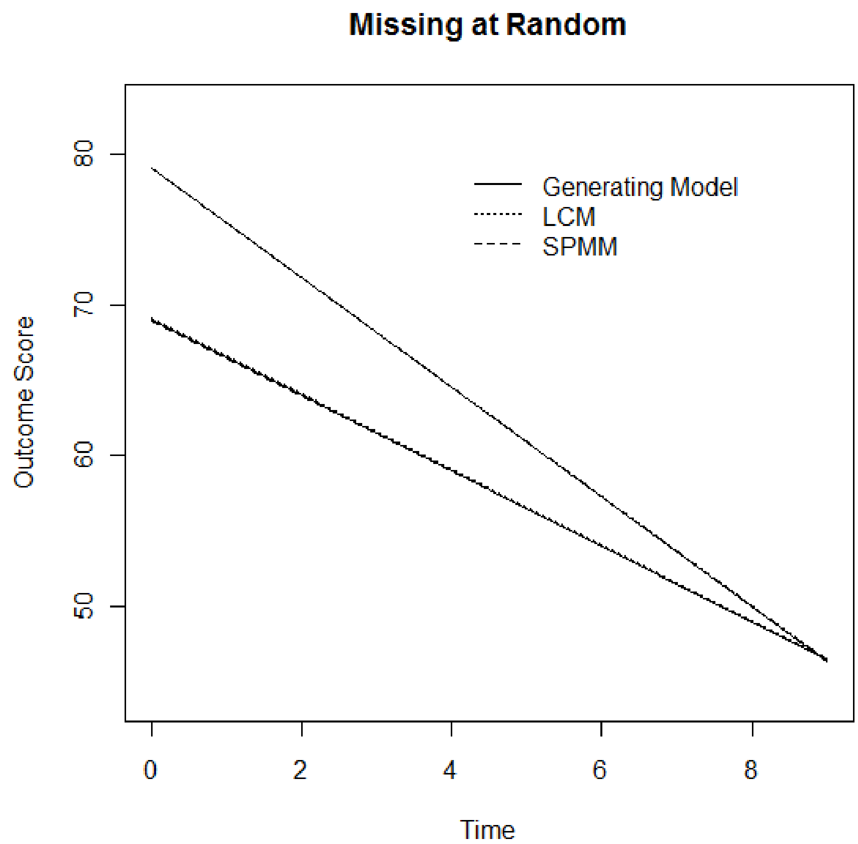


Figure 4. Comparison of LCM- and Binary SPMM-Implied Trajectories for $x_i = 0$ and $x_i = 1$ when the 'missing at random' assumption is met; results from best AIC are shown here

Table 2 illustrates that both the LCM and the SPMM produce fixed effect and variance component estimates with little bias (recall that SB is considered acceptable within the range of ± 40%), The RMSE values presented in Table 2 also indicates that LCM is slightly more efficient in recovering variance components than the SPMM, but that efficiency is about equivalent for fixed effect estimates. The finding that the LCM more efficiently recovers variance component estimates, but not fixed effects, might result from the fact that the SPMM misspecifies the random effects (e.g., Verbeke & Lesaffre, 1997; Litière, Alonso, & Molenberghs, 2008). Table 2 further indicates that there is no strong empirical reason to rely on the AIC or the BIC when the missing data mechanism is MAR.

Results for Hypotheses 2 and 3: Trajectory Recovery under MNAR

It was expected that the SPMM would recover trajectory estimates better than the LCM when the missing data mechanism was random coefficient-dependent, but that neither model would recover trajectories well under an outcome dependent MNAR process. Table 3 compares SB and RMSE values across MNAR study conditions and models, and Figure 5 shows the average LCM and Binary SPMM performance under the four MNAR conditions.

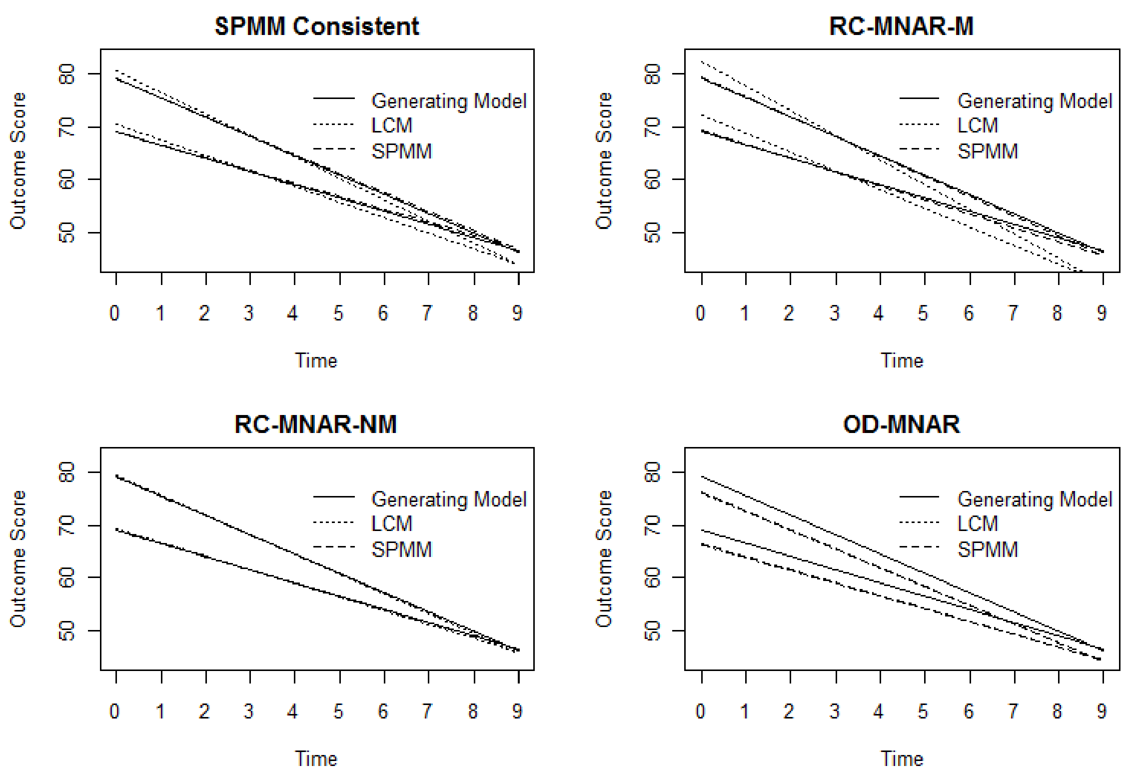


Figure 5. Comparison of LCM- and Binary SPMM-Implied Trajectories for $x_i = 0$ and $x_i = 1$ under a variety of non-random missing data mechanisms: SPMM consistent (top left), RC-MNAR-M (top right), RC-MNAR-NM (bottom left), and OD-MNAR (bottom right); results from best AIC are shown here

Beginning with the condition most favorable to the SPMM relative to the LCM (SPMM-consistent missingness), Table 3 shows that LCM-implied fixed effect estimates of the growth factor means are substantially biased, but that regression parameters are relatively unbiased, whereas SPMM-implied fixed effect estimates are all within the acceptable range for SB. Additionally, RMSE values are also moderately lower for the SPMM-implied fixed effect estimates of growth factor means. Except for estimated variation in the random slope, the LCM-implied variance component estimates are within the acceptable bias range. The SPMM-implied variance component estimates are all relatively unbiased, and the RMSE is moderately lower for the SPMM-implied parameters than for the LCM-implied parameters.

Moving to the RC-MNAR-M condition, the next most favorable condition for the SPMM, the same pattern of results is observed for the LCM (i.e., growth factor means and variance component estimates are substantially biased but regression parameters are unbiased). Comparing these results with the SPMM-implied estimates shows that SPMM-implied fixed effect and variance component estimates are substantially less biased than the estimates implied by the LCM. Indeed, bias in SPMM-generated estimates reach an ‘acceptable’ level of bias for almost all parameter estimates (the exception being a marginally unacceptable level of bias in the random slope variance when the AIC is used for class enumeration). However, the RMSE of the random slope variance and the covariance between the random intercept and random slope is more efficient under the LCM.

Moving next to the RC-MNAR-NM condition, Table 3 shows that the brunt of the bias induced by this missingness mechanism lies in the variance component estimates, rather than in the fixed effects. This is expected since the RC-MNAR-NM removes cases from either tail of the random slope distribution, leaving the mean relatively unchanged but substantially reducing the observed population variability. In this condition, bias in the SPMM-implied fixed effect estimates and variance component estimates are both lower than the LCM-implied estimates, but SPMM-implied variance component estimates never reach an acceptable level of bias.

Finally, as expected, fixed effect estimates for the intercept are substantially biased, regardless of whether the LCM or SPMM is used under an OD-MNAR missing data process. Variance component estimates are also biased under OD-MNAR, and SPMM is not useful for correcting these. In this case, RMSE values suggest that LCM performs better than the SPMM because the estimates are less variable.

It is important to note that a researcher who is not privy to the process underlying the missing data would not be able to distinguish between an OD-MNAR process and a MAR process because both would result in similar parameter estimates under LCM and SPMM.

Table 3. *Bias and Efficiency of Trajectory Recovery under Several MNAR Mechanisms*

	LCM		SPMM (Best AIC)		SPMM (Best BIC)	
	SB (%)	RMSE	SB (%)	RMSE	SB (%)	RMSE
SPMM-Consistent						
Fixed Effects						
Conditional Intercept (α_0)	**76.84**	2.4	1.60	1.88	10.53	1.90
Conditional Slope (α_1)	**-128.57**	.57	11.43	.35	-5.56	.36
Intercept Predictor (γ_0)	-1.85	2.70	-1.14	2.73	2.63	2.76
Slope Predictor (γ_1)	6.25	.48	0	.45	.44	.45
Variance Components						
Residual Intercept Variance (ψ_{00})	-2.87	43.59	-5.88	44.76	-6.68	44.04
Residual Slope Variance (ψ_{11})	**-60.81**	5.74	-2.48	1.26	-19.02	1.30
Covariance (ψ_{01})	27.79	6.41	2.45	5.10	10.99	5.09
RC-MNAR-M						
Fixed Effects						
Conditional Intercept (α_0)	**163.16**	3.63	15.57	2.22	10.91	2.14
Conditional Slope (α_1)	**-404.00**	1.05	-37.84	.41	-25.64	.40
Intercept Predictor (γ_0)	6.92	2.61	8.33	2.66	7.49	2.62
Slope Predictor (γ_1)	-2.70	.37	-2.00	.50	0	.47
Variance Components						
Residual Intercept Variance (ψ_{00})	-21.22	42.26	-6.81	50.82	.21	47.95
Residual Slope Variance (ψ_{11})	**-335.80**	2.84	**-40.54**	6.23	-18.49	5.93
Covariance (ψ_{01})	**129.86**	6.92	8.72	9.84	-1.00	9.42

	LCM		SPMM (Best AIC)		SPMM (Best BIC)	
	SB (%)	RMSE	SB (%)	RMSE	SB (%)	RMSE
RC-MNAR-NM						
Fixed Effects						
Conditional Intercept (α_0)	22.60	1.81	2.76	1.81	2.76	1.81
Conditional Slope (α_1)	-37.50	.26	-3.57	.28	-3.57	.28
Intercept Predictor (γ_0)	3.28	2.44	5.20	2.48	4.08	2.44
Slope Predictor (γ_1)	-8.82	.34	-9.09	.34	-9.09	.34
Variance Components						
Residual Intercept Variance (ψ_{00})	-29.54	38.41	-24.22	50.58	-24.96	41.37
Residual Slope Variance (ψ_{11})	**-270.51**	2.25	**-118.97**	3.74	**-166.34**	2.69
Covariance (ψ_{01})	**130.15**	6.52	**71.03**	8.10	**90.70**	6.96
OD-MNAR						
Fixed Effects						
Conditional Intercept (α_0)	**-152.78**	3.29	**-126.29**	3.12	**-138.50**	3.19
Conditional Slope (α_1)	28.00	.26	23.08	.27	32.00	.37
Intercept Predictor (γ_0)	-15.66	2.52	-12.41	2.79	-12.88	2.64
Slope Predictor (γ_1)	8.33	.36	18.42	.39	10.81	.37
Variance Components						
Residual Intercept Variance (ψ_{00})	**-85.80**	52.03	**-79.56**	188.30	**-84.93**	144.71
Residual Slope Variance (ψ_{11})	**-55.17**	.99	**-51.69**	2.23	**-53.41**	1.49
Covariance (ψ_{01})	14.87	4.64	19.41	8.39	15.37	7.97

Note. Standardized bias (SB) values above 40% or below -40% are bolded to indicate severe bias

It was expected that bias in variance components might lead to bias in the standard errors of the fixed effects. Standard error bias of the fixed effects, which is presented as a ratio of the the standard error estimates (where a ratio of one means that the estimates are unbiased) to empirical standard deviation of the point estimates, is presented in Table 4. As a baseline measure, the average standard errors estimated using LCM when the missing data mechanism is MAR is slightly lower than the empirical standard deviations of the sampling distributions for the fixed effect point estimates. Compared with LCM results, the SPMM-implied ratio of average standard error estimates to the empirical standard deviations of the sampling distributions for the four fixed effects parameters are approximately equivalent under all of the missing data mechanisms. However, the ratio of estimated standard errors tends to be empirical standard errors is on the small side (indicating increased risk for Type I errors) when the Binary SPMM is used.

In other words, the Binary SPMM (but not the Summary SPMM) leads to false confidence in the aggregate growth parameter values. This only occurs when the SPMM is used as an approximation, and not when it is literally true (i.e., MAR or SPMM-consistent). Perhaps the summary indicator is a more reliable measure of missingness than the binary indicators, leading to more precise estimates.

Table 4. *Comparison of Average Standard Error Estimates and Empirical Standard Deviation of Sampling Distributions for Fixed Effect Parameters by Missingness Condition and Model*

	LCM			Binary SPMM			Summary SPMM		
	Average $\overline{SE}_j$	Empirical SD	Ratio	Average $\overline{SE}_j$	Empirical SD	Ratio	Average $\overline{SE}_j$	Empirical SD	Ratio
MAR									
α_0	1.74	1.84	.94	1.73	1.83	.94	1.74	1.87	.93
α_1	.30	.33	.91	.30	.35	.86	.31	.34	.91
γ_0	2.46	2.65	.93	2.46	2.64	.93	2.47	2.66	.93
γ_1	.43	.45	.96	.43	.48	.90	.43	.50	.86
SPMM-Consistent									
α_0	1.85	1.90	.97	1.87	1.89	.99	2.03	1.90	1.07
α_1	.35	.35	1.00	.33	.36	.92	.36	.33	1.09
γ_0	2.61	2.71	.96	2.60	2.65	.98	2.63	2.62	1.00
γ_1	.49	.48	1.02	.42	.45	.93	.42	.41	1.02
RC-MNAR-M									
α_0	1.88	1.90	.99	1.97	2.08	.95	2.03	2.11	.96
α_1	.25	.25	1.00	.28	.36	.78	.30	.32	.94
γ_0	2.66	2.60	1.02	2.63	2.63	1.00	2.63	2.67	.99
γ_1	.36	.37	.97	.33	.48	.69	.31	.36	.86
RC-MNAR-NM									
α_0	1.78	1.77	1.01	1.78	1.82	.98	1.78	1.81	.98
α_1	.24	.24	1.00	.25	.28	.89	.25	.28	.89
γ_0	2.51	2.44	1.03	2.51	2.45	1.02	2.51	2.46	1.02
γ_1	.35	.34	1.03	.34	.33	1.03	.33	.33	1.00

	LCM			Binary SPMM			Summary SPMM		
	Average $\overline{SE}_j$	Empirical SD	Ratio	Average $\overline{SE}_j$	Empirical SD	Ratio	Average $\overline{SE}_j$	Empirical SD	Ratio
OD-MNAR									
α_0	1.76	1.80	.98	1.74	1.87	.93	1.84	1.79	1.03
α_1	1.74	1.90	.92	.27	.25	1.08	.26	.27	.96
γ_0	2.50	2.49	1.00	2.44	2.64	.92	2.44	2.65	.92
γ_1	.38	.36	1.06	.38	.37	1.03	.38	.37	1.03

Note. SPMM results are based on the solution with the lowest BIC for each replication.

Results for Hypotheses 4 and 5: Summary SPMM versus Binary SPMM

It was hypothesized that the Binary SPMM would be better able to accommodate a non-monotonic mechanism than the Summary SPMM, but that both models would be equivalently unbiased (or biased) in all other conditions. It was also hypothesized that the Summary SPMM would be more efficient than the Binary SPMM when both equivalently capture information about the missing data process (i.e., under a MAR mechanism, a RC-MNAR-M, and a SPMM-consistent mechanism).

SB and RMSE bias of the Binary and Summary SPMM-implied estimates are reported in Table 5, and Figure 6 compares the Binary and Summary SPMMs across the four MNAR conditions (trajectories implied under a MAR mechanism were on top of the generating lines). Results show support for Hypotheses 4; there are no meaningful or consistent differences across the two models with respect to parameter bias. Both models recover all parameters well under MAR and SPMM-consistent mechanisms, both recover parameters adequately well under an RC-MNAR-M mechanism, both struggle with variance component recovery under RC-MNAR-NM, and both produce quite biased parameter estimates under an OD-MNAR mechanism. It was expected that the Summary SPMM might provide somewhat more biased estimates under a RC-MNAR-NM mechanism when compared with the Binary SPMM, but this was not the case. This result is probably due to the fact that the Binary SPMM does a poor job at recovering variance component estimates, and is not an improvement over the LCM; thus, the Summary SPMM does not perform any worse (or better) than the Binary SPMM.

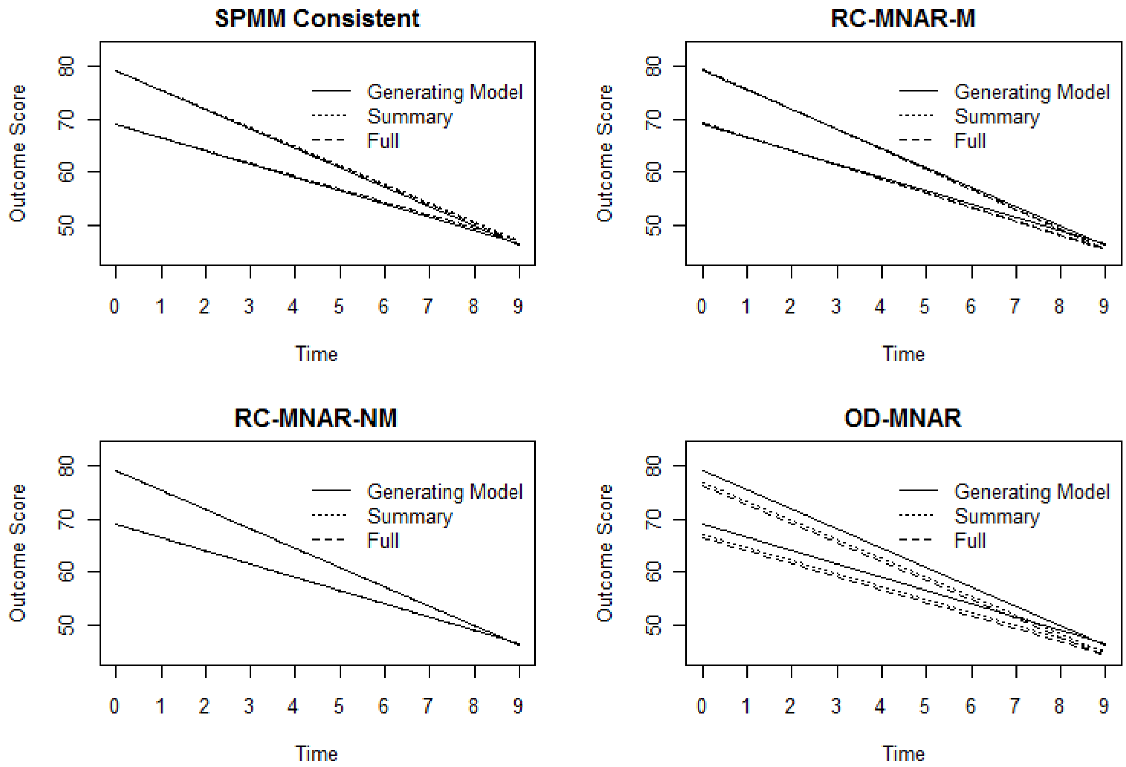


Figure 6. Comparison of and Binary SPMM- (i.e., "Full") and Summary SMM- Implied Trajectories for $x_i = 0$ and $x_i = 1$ under a variety of non-random missing data mechanisms: SPMM consistent (top left), RC-MNAR-M (top right), RC-MNAR-NM (bottom left), and OD-MNAR (bottom right); results from best AIC are shown here

Table 5. *Standardized Bias and Root Mean Squared Error of Binary and Summary SPMM- Implied Parameter Estimates*

	Binary SPMM				Summary SPMM			
	Best AIC		Best BIC		Best AIC		Best BIC	
	SB (%)	RMSE	SB (%)	RMSE	SB (%)	RMSE	SB (%)	RMSE
MAR								
α_0	.55	1.82	0	1.83	-3.26	1.84	-1.64	1.83
α_1	0	.35	2.86	.35	-2.94	.34	-2.94	0.34
γ_0	0	2.68	.75	2.71	9.70	2.70	14.13	2.67
γ_1	4.26	.47	2.13	.48	0.00	.48	0.00	.48
ψ_{00}	-16.49	56.01	-15.84	49.22	-18.14	63.38	-18.48	61.44
ψ_{11}	-17.76	1.69	-16.67	1.50	-25.45	1.62	-25.69	1.58
ψ_{01}	10.73	6.68	8.98	6.09	13.54	6.81	9.90	6.54
SPMM-Consistent								
α_0	1.60	1.88	10.53	1.90	8.25	1.95	7.73	1.95
α_1	11.43	.35	-5.56	.36	17.65	.34	17.65	0.34
γ_0	-1.14	2.73	2.63	2.76	-8.89	2.75	-8.89	2.75
γ_1	0	.45	.44	.45	9.30	.42	7.14	.42
ψ_{00}	-5.88	44.76	-6.68	44.04	-4.66	45.22	-3.97	45.19
ψ_{11}	-2.48	1.26	-19.02	1.30	1.85	1.10	1.85	1.10
ψ_{01}	2.45	5.10	10.99	5.09	-4.07	5.23	-3.30	5.20

	Binary SPMM				Summary SPMM			
	Best AIC		Best BIC		Best AIC		Best BIC	
	SB (%)	RMSE	SB (%)	RMSE	SB (%)	RMSE	SB (%)	RMSE
RC-MNAR-M								
α_0	15.57	2.22	10.91	2.14	2.40	2.09	0.48	2.09
α_1	-37.84	.41	-25.64	.40	-18.18	.33	-18.18	0.33
γ_0	8.33	2.66	7.49	2.62	5.56	2.60	4.87	2.60
γ_1	-2.00	.50	0	.47	5.26	.35	5.26	.35
ψ_{00}	-6.81	50.82	.21	47.95	-20.90	47.72	-21.16	47.78
ψ_{11}	**-40.54**	6.23	-18.49	5.93	-33.53	5.38	-34.50	5.37
ψ_{01}	8.72	9.84	-1.00	9.42	14.73	8.90	14.75	8.90
RC-MNAR-NM								
α_0	2.76	1.81	2.76	1.81	8.29	1.81	8.84	1.81
α_1	-3.57	.28	-3.57	.28	-3.57	.28	-7.14	0.28
γ_0	5.20	2.48	4.08	2.44	9.64	2.45	9.64	2.44
γ_1	-9.09	.34	-9.09	.34	3.03	.33	3.03	.33
ψ_{00}	-24.22	50.58	-24.96	41.37	-26.52	60.96	-26.71	53.33
ψ_{11}	**-118.97**	3.74	**-166.34**	2.69	**-108.13**	3.74	**-113.82**	3.52
ψ_{01}	**71.03**	8.10	**90.70**	6.96	**75.11**	8.36	**76.43**	7.76
OD-MNAR								
α_0	**-126.29**	3.12	**-138.50**	3.19	**-118.13**	2.86	**-117.49**	2.87
α_1	23.08	.27	32.00	.37	23.08	.27	26.92	0.27
γ_0	-12.41	2.79	-12.88	2.64	-15.33	2.75	-15.56	2.73
γ_1	18.42	.39	10.81	.37	2.78	.37	5.56	.37

	Binary SPMM				Summary SPMM			
	Best AIC		Best BIC		Best AIC		Best BIC	
	SB (%)	RMSE	SB (%)	RMSE	SB (%)	RMSE	SB (%)	RMSE
ψ_{00}	**-79.56**	188.30	**-84.93**	144.71	**-60.43**	114.56	**-61.34**	86.63
ψ_{11}	**-51.69**	2.23	**-53.41**	1.49	**-52.27**	1.29	**-51.72**	1.19
ψ_{01}	19.41	8.39	15.37	7.97	18.46	5.81	18.62	5.17

Note. α_0 is the average growth intercept; α_1 is the average slope of growth; γ_0 is the effect of x_i on the random intercept; γ_1 is the effect of x_i on the random slope; ψ_{00} is the variance of the random intercept; ψ_{11} is the variance of the random slope; ψ_{01} is the covariance of the random intercept and slope; SB bias above 40% or below –40% is bolded.

It was hypothesized that the Summary SPMM would provide more efficient estimates than the Binary SPMM. Efficiency was operationalized using the RMSE. Table 5 shows that this hypothesis is not generally true. When the missing data mechanism is MAR or SPMM-consistent, both models are about equally efficient. However, when the missing data mechanism is RC-MNAR-M or OD-MNAR, the Summary SPMM is more efficient than the Binary SPMM. When the missing data mechanism is RC-MNAR-NM, both models are equally efficient for recovery of the fixed effects, but the Binary model recovers variance component estimates more efficiently. This effect is most prominent when the BIC is used for class enumeration, suggesting that fewer classes are better for recovering variance components more efficiently (at least when the model is not sufficient for approximating the missing data mechanism well).

In sum, both models perform equally well when the SPMM is the 'true' model (i.e., under MAR or SPMM-consistent MNAR missingness), and the Summary SPMM is more efficient than the Binary SPMM when the SPMM provides some approximation of the missing data mechanism, whether it is a good approximation (i.e., RC-MNAR-M), or a bad approximation (i.e., OD-MNAR).

Summary and Discussion of Study 1

The goal of Study 1 was to test relative performance of the LCM, the Binary SPMM, and the Summary SPMM under a variety of realistic data conditions that might be encountered when analyzing longitudinal data. The first point to take away from Study 1 is the SPMM does not solve the problem of missing data. In the words of

Demirtas and Schafer (2003), "the best way to handle missing values is to not have them" (pp. 2573). Barring that possibility, Study 1 makes it clear that, when random coefficient-dependent missingness is suspected, the SPMM is a helpful tool in testing the sensitivity of the MAR assumption that is implied by the LCM.

Results from Study 1 show that the LCM, which assumes that missing data are MAR, produces biased estimates of growth factor fixed effects and variances whenever the MAR assumption is violated. As a caveat, Study 1 also showed that regression effects are robust to violations of the MAR assumption, at least for the conditions tested here.[10] Study 1 showed that the SPMM, whether Binary or Summary, produces badly biased fixed effect estimates only when the missing data mechanism is OD-MNAR (i.e., a time-varying process). Further, the SPMM is not able to recover variance components well when the missing data mechanism is RC-MNAR-NM. The SPMM provides improved parameter estimates over the LCM when a non-ignorable random coefficient-dependent missing data process is present, if the process is monotone or discrete. These results were generally expected given that SPMMs are designed specifically to accommodate random coefficient dependent missing data processes.

Under no condition did the SPMM provide worse parameter estimates than the LCM; however, variance component estimates were less efficient when the SPMM was used with MAR missingness. In other words, there is no serious harm done when the SPMM is used instead of the LCM, even when the data are MAR. However, a researcher who obtains effectively identical point estimates when comparing results obtained using

[10] It is possible to imagine an MNAR scenario in which the regression coefficients might be biased. For example, the regression coefficients might be biased if the severity of non-random selection depends on the level of x_i .

an LCM with results obtained using a SPMM should rely on LCM results for the sake of parsimony and efficiency.

The Summary SPMM was shown to perform as well as the Binary SPMM under all conditions, and results suggested that the Summary SPMM provides slightly more efficient estimates than the Binary SPMM when the missing data process is monotonically random coefficient dependent, but it not exactly consistent with the model (i.e., RC-MNAR-M).

Limitations and Future Directions

Study 1 assessed relative SPMM and LCM performance under only five of many possible missing data conditions. While the five conditions tested represent a broad range of conditions that would possibly be encountered with longitudinal social science data, this simulation study was necessarily limited in its generalizability. Most notably absent from the conditions tested in Study 1 were more asymmetric nonmonotone random coefficient dependent mechanisms, conditions containing multiple mechanisms operating simultaneously, and conditions involving the regression parameters. In spite of this limitation, the results here are informative, at least for predicting what might happen in real world data conditions. Real world situations involving multiple missing data mechanisms would be prey to multiple types of parameter bias, and more complex missing data mechanisms that involved multiple growth factors would contain bias in more of the parameters.

A more pressing limitation of Study 1 is its failure to consider monotone dropout, a commonly observed type of missingness in longitudinal research. Study 2 will compare dropout with the erratic mechanism studied in this chapter, in addition to examining the role that observation length and proportion of missing data play in LCM and SPMM performance.

Chapter 3

STUDY 2: A CLOSER LOOK AT SPMM PERFORMANCE ACROSS DATA CONDITIONS

Study 1 compared LCM and SPMM performance across a range of missing data mechanisms; however, it was limited in the range of data conditions that were assessed. To further investigate LCM and SPMM performance across a range of data conditions, Study 2 crossed number of observation occasions (5, 10, or 20) with erratically spaced missing data or complete attrition, and with proportion of missing data (.30 or .60), resulting in a 3 x 2 x 2 full factorial design (i.e., 12 conditions). The missingness mechanism studied was the RC-MNAR-M mechanism from Study 1.[11] Because it was determined from Study 1 that the Summary SPMM performs better than the Binary SPMM under RC-MMNAR-M in the sense that Summary SPMM-generated estimates are more efficient than Binary SPMM-generated estimates and Summary SPMM-generated standard errors are generally unbiased, and because the Summary SPMM is faster to implement, only the Summary SPMM performance was compared with LCM performance in Study 2.

The following hypotheses were tested in Study 2:

1. The LCM-implied trajectories will be least efficient when growth factors are poorly determined (i.e., when there are fewer repeated measures due to few measurement

[11] RC-MNAR-M is the mechanism that is typically described when random coefficient dependent missingness is discussed in the missing data literature (e.g., Little, 2009).

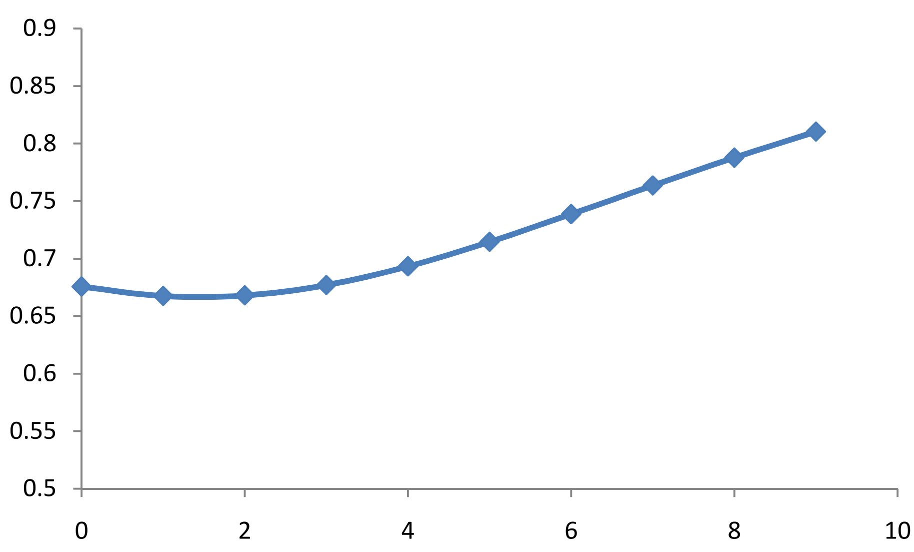
occasions or a higher proportion of missingness and when subjects drop out rather than providing erratic information across the span of measurement occasions). The hypothesis that growth factor parameter estimates will be less efficient when less information is available with which to determine the growth factors is not unique to the problem of non-randomly missing data; data conditions leading to higher factor determination should be the same whether missing data are randomly missing or not.

In addition to the efficiency hypothesis, however, it may be reasonable to expect that growth factor means and variance components will also be less biased when the factors are better determined. As the correlation between the observed information (i.e., the observed repeated measures) and the growth factor approaches unity, the random coefficients become less 'latent' and more determined. In turn, the random coefficient-dependent missingness mechanism should approach ignorability, resulting in less biased parameter estimates when ignorability is assumed.

The amount of available information (e.g., observation length) is directly linked to growth factor determinacy (i.e., the reliability of a factor score estimate based on the manifest variables, or the correlation between a factor score estimate and a true factor score, which can be measured as follows:

$$\rho^2 = \boldsymbol{\lambda}'\boldsymbol{\Sigma}_{\mathbf{y}}^{-1}\boldsymbol{\lambda}; \tag{3.1}$$

Brown, 1910; Spearman, 1910; Guttman, 1955; Mulaik, 1976; Green, 1976; Mulaik & McDonald, 1978; Bollen, 1980; Grice, 2001), where ρ^2 is the reliability, $\boldsymbol{\lambda}$ contains factor loadings, and $\boldsymbol{\Sigma}_y$ is the covariance matrix of the manifest variables. The relationship between the nature of missingness and growth factor determinacy / reliability is less obvious. The effects of the two processes (erratic missingness or dropout) on reliability would differ to the extent that the communality of observed measures changes systematically over time. In this instance, an increasing proportion of the variance in repeated measures is explained by individual differences in development (i.e., by the growth factors), and a decreasing proportion of the variance is due to residual variation. This is the case because the factor loading matrix for the slope coefficient is structured so that the intercept is located at the first observation occasion, with an increasing contribution of the slope for each repeated measure.[12] That is, a larger proportion of the variation in the items that have been observed is due to unique variance for the items that are observed than for the items that were not observed when a dropout mechanism is operating. Figure 7 shows the relationship between the proportion of variance explained in the items that were generated using the simulation values in this manuscript, as a function of time, up to $t = 9$.



[12] Variance explained by the random slope increases in magnitude as the factor loadings increase in magnitude, regardless of sign.

Figure 7. The relationship between time (on the x-axis) and the proportion of item variance explained (on the y-axis) by the growth factors in Study.

Since more of the item variance is noise at the early time points, it is expected that growth factors should be less determined when missing data are missing due to monotonic dropout than if missing data are erratically missing.[13]

In a sense, any effect of the nature of missingness that is observed in Study 2 thus depends on an increase in the contribution of the random slope on proportion of variance explained in the items. That is, it is assumed here that dropout results in the failure to observe items that are most informative about the growth process. The same effect might not be observed if random coefficient dependent missingness were, for instance, random intercept dependent.

2. The second main hypothesis of Study 2 is that SPMM may have difficulty recovering a sufficient number of classes to accommodate informative missingness when there are few observed repeated measures (i.e., when there are few measurement occasions or when there is a high proportion of missing data) and when the missingness mechanism is dropout rather than erratic. Particularly when only five repeated measures are observed and when 60% of the items are missing when ten or fewer repeated measures are observed for an individual, it may be impossible to extract four- or five-classes in the SPMM because there will not be enough information available with which to identify growth factor means for each class separately. Hedeker and Gibbons (1997) and Little (1993, 1994, 1995) have described difficulty with empirical identification of separate

[13] The relationship between item reliability and time is not linearly / monotonically increasing because there is a negative covariance between the growth factors in the data generating mechanism. The proportion of variance explained is equal to: $\frac{VAR(\eta_0)+\lambda^2_{1t}VAR(\eta_1)+2\lambda_{1t}COV(\eta_0,\eta_1)}{VAR(\eta_0)+\lambda^2_{1t}VAR(\eta_1)+2\lambda_{1t}COV(\eta_0,\eta_1)+VAR(\varepsilon)}$.

growth patterns in traditional PMM when an insufficient number of repeated measures are observed with which to identify a trajectory within certain classes. It is therefore expected that this situation will be ameliorated to some extent when latent classes are used in place of deterministic, observed patterns because individuals with longer observation lengths will contribute some information to shorter growth groups; however, the problem of empirical underidentification is not likely to be completely solved when many classes are extracted and there are few repeated measures per individual, or when observed repeated measures are located in close proximity with one another (i.e., with monotone dropout). On the opposite side of the observation length spectrum, Eggleston, Laub, and Sampson (2004; using a single dataset) found that an increased observation length was related to the extraction of more latent classes, so it is expected that having more information (i.e., 30% as opposed to 60% missingness, or 20 repeated measures), will be related to the extraction of additional classes.

In turn, it is expected that the extraction of additional classes will be linked to better SPMM performance due to an improved ability to establish conditional independence from individual growth trajectories and the missing data indicator.

Data Generation

The same population-generating model that was used in Study 1 was also used in Study 2, except that the number of repeated measures varied between five and 20. The proportion of variance in the repeated measures that was explained at the intercept was 68.87% across conditions, and the proportion of variance explained by the conditional model was 74.72% for the fifth repeated measure, 87.97% for the tenth repeated measure, and 96.49% for the twentieth repeated measure.

After generation of the complete data, observations were selectively deleted according to a RC-MNAR-M mechanism so that the overall probability of missingness was .30 or .60. The erratic RC-MNAR-M mechanism conditions were constructed so that the log odds of missingness for any given data point for an individual was a linear function of their random slope, varying the intercept of the linear function to achieve 30% or 60% missingness, as shown below[14]:

$$\begin{aligned}
&\text{30\% missing: } p(r_{it}=1\,|\,\eta_{1i}) = \frac{e^{-.85+2(\zeta_{1i}/\psi_{11})}}{1+e^{-.85+2(\zeta_{1i}/\psi_{11})}} \\
&\text{60\% missing: } p(r_{it}=1\,|\,\eta_{1i}) = \frac{e^{.40+2(\zeta_{1i}/\psi_{11})}}{1+e^{.40+2(\zeta_{1i}/\psi_{11})}} \ .
\end{aligned} \tag{3.2}$$

The dropout RC-MNAR-M mechanism conditions were constructed so that the first missing observation was expected to occur either 40% of the way through the study (leading to 60% missingness) or 70% of the way through the study (leading to 30% missingness) for an average individual, with dropout expected to occur earlier for individuals with higher random slopes. To maintain consistency with the erratic missingness condition with respect to severity of informativeness, the same regression coefficient was used to relate ζ_{1i} to the log odds of missingness. Intercept terms for the linear model varied by percent of missingness in order for the probability of missingness

[14] The severity of informativeness did not vary by proportion of missingness in the sense that coefficients linking ζ_{1i} with $p(r_{ti}=1)$ remain constant across conditions. More severe informativeness has predictable implications for bias of model parameters under LCM. The purpose of altering the proportion of missing data was to test the limit of the SPMM's capability to draw information regarding dependencies between missing data patterns and repeated measures when many observations are missing and to study the effect of proportion missingness on growth factor determinacy in LCM.

for a given occasion to reach an expected survival time (defined as the first occasion of missingness, because missingness is monotone in the dropout conditions) at the desired measurement occasion. Expected survival time and probability of missingness for each occasion are listed in Table 6.

Table 6. *Expected Survival Time and Probability of Missingness for* y_{ti}

	Expected Survival Time	Missingness Probability
Five Repeated Measures		
30% missing	3.5	.18
60% missing	2	.29
Ten Repeated Measures		
30% missing	7	.09
60% missing	4	.16
Twenty Repeated Measures		
30% missing	14	.05
60% missing	8	.08

Note. Expected survival time denotes the observation occasion at which there is a 50% cumulative probability of a missing observation (the intercept is located at $t = 0$). This corresponds with individual item missingness probabilities in the right-hand column.

Five hundred replications of sample size 300 were generated for each population condition.

Data Analysis

Data analysis in Study 2 matched the Study 1 procedure, except that only the Summary SPMM (one through five classes per replication)[15] and the LCM were applied to each dataset. When estimating the Summary SPMM, there were two options for constructing the summary indicator: time of last measurement occasion and number of observations. In the dropout condition, these alternative summaries would provide identical information. In the erratic missingness condition, the last observation occasion

[15] Only four classes were estimated in the conditions with five repeated measures because, in SPMM and related latent mixture modeling approaches, the number of latent classes K is less than the number of observed patterns of missingness. For the dropout conditions, this means that $K < 5$ in the conditions with only 5 observation occasions.

would not be as meaningful as the number of observations; thus, number of observations was used as a summary indicator for all conditions.

Within a replication, a solution was removed from possible AIC / BIC selection if it: 1) did not converge to a positive definite solution; 2) was a visual outlier, or 3) contained a class with $\hat{\pi}_k < .10$. These were the same criteria that were used in Study 1.

Raw bias (RB), standardized bias (SB) and root mean squared error (RMSE) were calculated for each study condition, along with coverage rates and class enumeration based on the AIC and BIC.

Results for Hypothesis 1: Higher Factor Score Determinacy is Related to Less Biased and More Efficient LCM Estimates in the Presence of Non-Randomly Missing Data

It was hypothesized that circumstances leading to increased growth factor determinacy would decrease the severity of the effects of random coefficient dependent missingness on parameter estimates when an LCM is used for data analysis, controlling for the strength of the association between the random coefficient and the probability of missingness. Specifically, more observations (resulting from more repeated measures or from a smaller proportion of missing data) were expected to increase the accuracy and precision of LCM-based trajectory estimates in the presence of random coefficient dependence missingness. Further, the nature of the pattern of missingness was expected to affect factor score determinacy; controlling for the proportion of missing data and the magnitude of the association between the random coefficient and the probability of missingness, erratically spaced missingness was expected to be related to higher factor score determinacy (and therefore more accuracy and precision) than monotonic missingness (i.e., dropout).

Figures 8 and 9 show RB, SB, and RMSE values, respectively, for the random slope mean (Figure 8) and variance (Figure 9) by percent of missingness (left) and nature

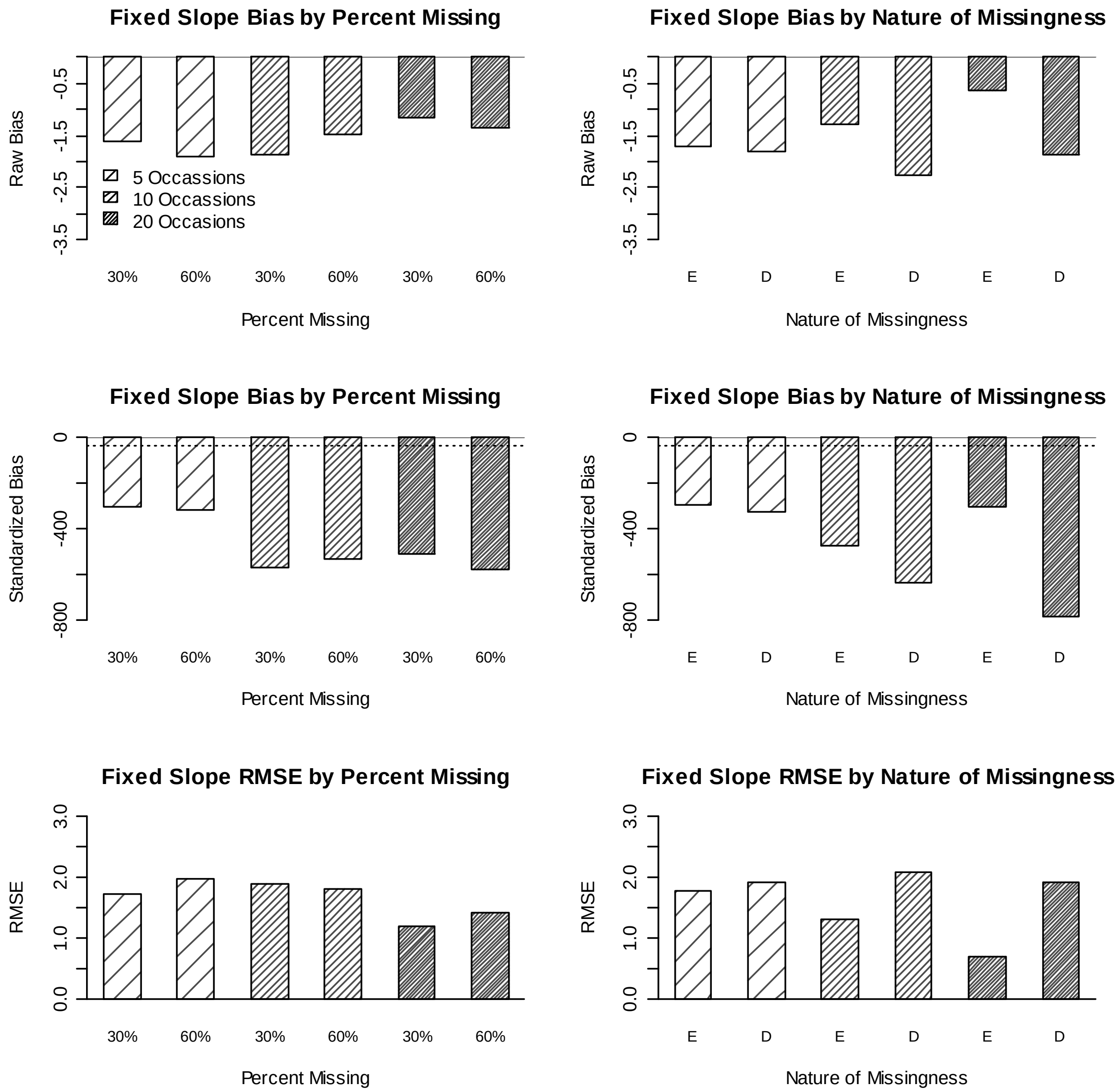


Figure 8. Raw bias (top), standardized bias (middle), and RMSE (bottom) of LCM-implied fixed effect estimates for the slope by percent of missingness (30% or 60%; left) and nature of missingness (erratic (E) or dropout (D); right) and by number of observation occasions (5, 10, or 20). A horizontal reference line is drawn at SB = -40 to indicate the cutoff for “acceptable” bias.

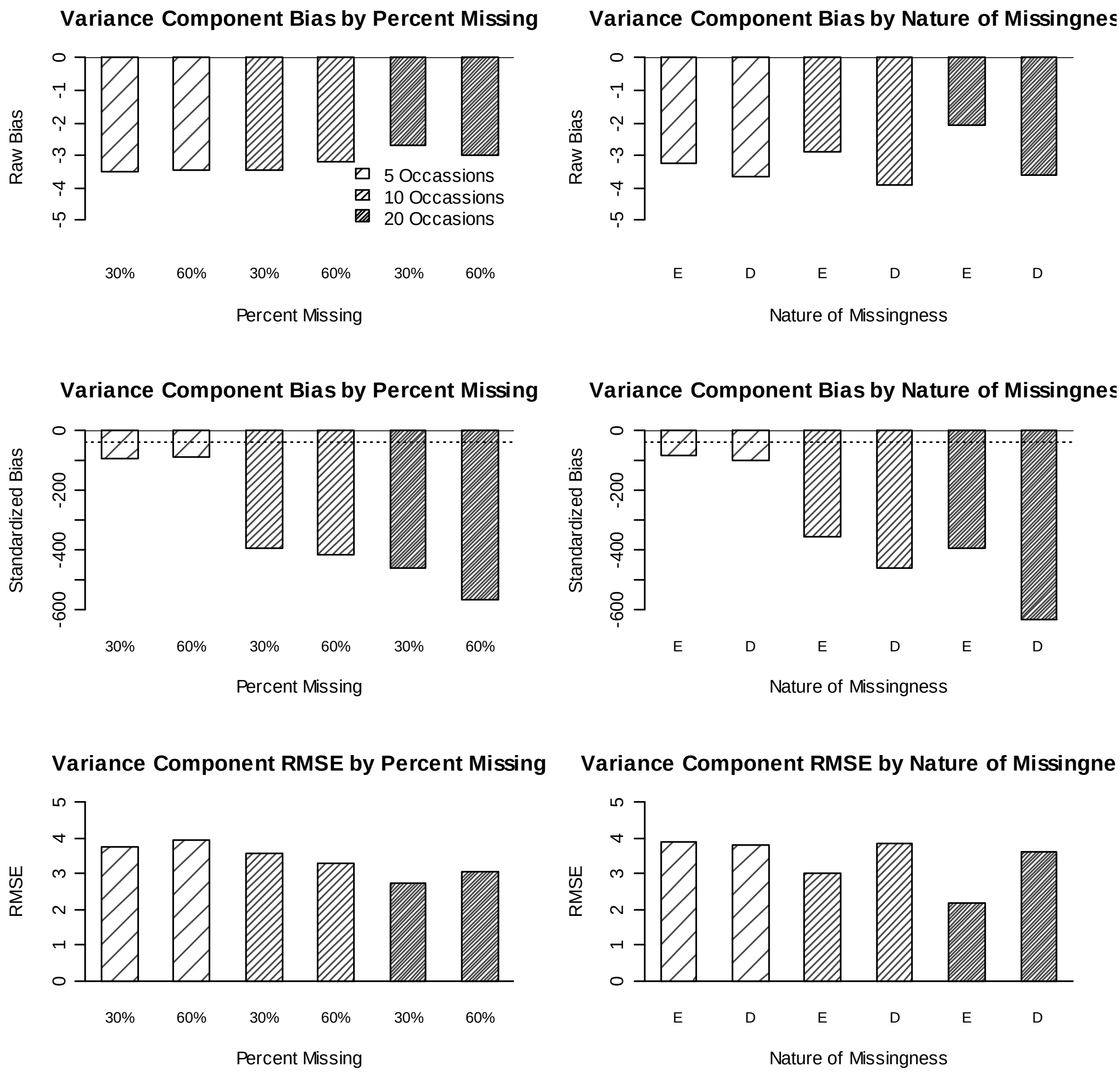


Figure 9. Raw bias (top), standardized bias (middle), and RMSE (bottom) of LCM-implied variance component estimates for the slope by percent of missingness (30% or 60%; left) and nature of missingness (erratic (E) or dropout (D); right) and by number of observation occasions (5, 10, or 20). A horizontal reference line is drawn at SB = -40 to indicate the cutoff for “acceptable” bias.

of missingness (i.e., erratic missingness or dropout; right), conditional on the number of observation occasions (each cell).[16]

Effect of Number of Observation Occasions. It was hypothesized that more observation occasions would lead to less biased and more efficient parameter estimates due to increased factor determinacy. The top panel of Figure 8 (RB) suggest a mild effect of the number of observation occasions on fixed effect bias in the expected direction, and the same is true for the variance shown in Figure 9. This effect appears to be reversed for the SB, but this is an artifact due to the decrease in standard errors resulting from more observations. The RMSE, shown in the bottom row of both Figures, concurs with the hypothesis that having more repeated measures leads to more precise estimates of growth (with respect to both mean change and population variation around the mean).

Effect of Percent of Missingness. It was hypothesized that growth factors would be better determined with more observed information, and thus that parameters related to growth would be more biased and less efficient as the percentage of missingness increased, even controlling for severity of the missingness mechanism. However, the left-hand panels of Figures 8 and 9 suggest that there is no substantial effect of the percent of missing data on parameter bias or RMSE. Hindsight suggests that the random subtraction of missing data should not lead to increased parameter bias, although the deletion of additional data *should* lead to less efficient parameter estimates. An inspection of the empirical standard deviations of the LCM-based estimates, shown in Table 7, suggests that the effect of proportion of missing data on efficiency depends on

[16] Because the random coefficient dependent mechanism was solely dependent upon the random slope, emphasis is placed on recovery of the fixed effect estimate of the rate change over time, and on the recovery of the variance of the random slope. These were the parameters that were affected in Study 1.

observation length and on the nature of missingness. For erratically missing data, a higher proportion of missing data is linked to larger empirical standard deviations (and standard error estimate) only when five repeated measures are present. For dropout, more missing data is linked to *lower* empirical standard deviations when five repeated measures are present. There is no effect of proportion of missing data as the number of repeated measures increases.

Table 7 also shows that the efficiency of variance component estimates is dramatically improved as the observation length increases; fixed effect estimates are more robust to study length.

Table 7. *Comparison of Empirical Standard Deviations for LCM-Based Slope Estimates*

	30% Missing		60% Missing	
	Dropout	Erratic	Dropout	Erratic
Five Observations				
Fixed Slope	.57	.48	.54	.65
Slope Variance	3.60	3.59	3.44	4.37
Ten Observations				
Fixed Slope	.36	.25	.27	.28
Slope Variance	.94	.78	.69	.85
Twenty Observations				
Fixed Slope	.23	.21	.24	.21
Slope Variance	.57	.57	.55	.50

Nature of Missingness. It was expected that the LCM would have more trouble accommodating missingness due to dropout than erratic missingness. This hypothesis was supported for both the fixed effect estimate and for the variance component, except for the cells with a high proportion of missing data and five repeated measures. In this case, dropout produces estimates that are about equally biased, but that have smaller standard errors.

Although results are largely consistent with the hypothesis regarding trends in bias as a function of growth factor determinacy, the fixed effect estimates generated by the LCM are severely biased under the MNAR mechanism imposed in this study, regardless of study condition. These results suggest that it is not safe to assume that a MNAR process can be ignored, even when study conditions are favorable for increasing growth factor score determinacy.

Both the fixed effect (Figure 8) and variance component estimates (Figure 9) are always downwardly biased (in the sense that growth trajectories are estimated to be decreasing more steeply than they actually are). The fixed effect estimates are downwardly biased because higher levels of the random slope were related to an increased probability of missingness. Variance components are downwardly biased because the removal of observations leads to less observed variation in the population.

Another general point to observe in Figures 8 and 9 is that trends in RB and SB are not always aligned. The SB of the point estimates is scaled by the standard error of the estimates; thus, it is not necessarily meaningful to compare SB across estimates derived from conditions with different lengths of observation. This is because SB will be *larger* as the number of observation occasions increases, simply because standard errors decrease as the number of observations increase. When comparing results across observation lengths, it may be more appropriate to compare RB and RMSE.

Results for Hypothesis 2: SPMM Should Perform Worse when Fewer Classes are Supported

Tables 8 – 10 show rates of convergence to a proper solution and case removal due to low class probability estimates ($\hat{\pi}_k$) or due to being a visual outlier for conditions with five, 10, and 20 observation occasions, respectively. These tables also include information about AIC- and BIC-based model selection. It is difficult to clearly express results of a study with 12 separate conditions. In an attempt to maximize lucidity, results are discussed in the following order: first, effects of observation length will be discussed. Next, effects of nature of missingness will be discussed conditional on occasion length. Finally, effects of proportion of missing data will be discussed conditional on occasion length.

It was expected that more classes would be supported as the observation length increased. This hypothesis was met in the sense that the *range* of classes that could be estimated was limited for the five repeated measure conditions. On the other hand, the same *average* number of classes were selected by the AIC and BIC across all observation length conditions.

It was also expected that more classes would be supported and extracted with an erratic missing data mechanism than with monotone dropout. This hypothesis was supported in the same sense that the hypothesis about the effect of the observation length was supported; more classes converged to proper, reasonable solutions with erratic missingness than with dropout, but only slightly more classes were chosen by the information criteria when the missing data mechanism was erratic. There was an interaction between the number of observation occasions and the nature of missingness with respect to class extraction. Regardless of whether the missingness mechanism was dropout or erratic, it was difficult for a SPMM to support more than two or three classes when only five repeated measures were present. As the number of repeated measures increased, more classes were supported with erratic missingness, but the SPMM still has difficult extracting many classes with dropout.

Table 8. *Rates of Convergence to a Proper Solution, Solution Deletion, and AIC and BIC Model Selection with 5 Observations*

	Monotone Dropout						Erratic Missingness					
Classes	Converged	Low π_k	Outlier	Remain	AIC	BIC	Converged	Low π_k	Outlier	Remain	AIC	BIC
30% Missing												
1	408	NA	0	408	2	2	433	NA	0	433	11	12
2	407	0	0	407	38	38	412	0	0	412	23	23
3	404	15	20	369	367	367	383	1	0	383	283	303
4	387	384	0	3	2	2	278	42	8	228	143	122
60% Missing												
1	416	NA	1	415	24	25	416	NA	2	414	24	25
2	383	0	1	381	13	15	382	0	1	381	13	15
3	370	0	1	369	248	254	370	0	1	369	248	254
4	351	17	12	322	185	176	351	17	13	321	185	176

Table 9. *Rates of Convergence to a Proper Solution, Solution Deletion, and AIC and BIC Model Selection with 10 Observations*

	Monotone Dropout						Erratic Missingness					
Classes	Converged	Low π_k	Outlier	Remain	AIC	BIC	Converged	Low π_k	Outlier	Remain	AIC	BIC
30% Missing												
1	500	NA	0	500	0	0	500	NA	0	500	0	0
2	500	0	1	499	78	79	500	0	0	500	3	3
3	499	17	4	468	356	356	498	1	2	495	142	153
4	496	402	3	91	42	41	485	10	6	469	317	309
5	487	432	2	53	24	24	428	318	2	108	38	35
60% Missing												
1	500	NA	0	500	0	0	500	NA	0	500	0	0
2	500	0	0	500	125	143	500	0	0	500	11	11
3	500	63	2	435	299	281	487	0	1	486	42	161
4	462	344	20	98	37	37	476	13	10	453	396	313
5	473	427	2	44	39	39	427	254	0	173	51	15

Table 10. *Rates of Convergence to a Proper Solution, Solution Deletion, and AIC and BIC Model Selection with 20 Observations*

	Monotone Dropout						Erratic Missingness					
Classes	Converged	Low π_k	Outlier	Remain	AIC	BIC	Converged	Low π_k	Outlier	Remain	AIC	BIC
30% Missing												
1	500	NA	0	500	0	0	500	NA	0	500	0	0
2	500	0	1	499	13	13	500	0	0	500	0	0
3	500	8	5	487	384	389	500	0	1	499	31	31
4	488	360	0	128	59	61	498	10	5	483	203	205
5	499	428	0	71	44	37	500	200	3	297	266	264
60% Missing												
1	500	NA	0	500	0	0	500	NA	1	499	0	0
2	500	0	0	500	11	11	500	0	1	499	5	5
3	500	7	6	487	372	378	500	4	1	495	106	107
4	485	346	1	138	70	70	493	72	13	408	195	198
5	500	433	0	67	47	41	500	185	2	313	194	190

Finally, it was expected that more classes would be supported with 30% missing data than with 60% missing data. While the same average number of classes were extracted across these conditions, *more* classes are supported with 60% missingness in the five repeated measure conditions (particularly when dropout was present). This trend is not as apparent when ten or twenty repeated measures are observed. This is an unexpected finding, and might possibly be explained if class extraction in the 60% dropout condition with five repeated measures purely reflects level differences in the intercept, rather than differences in the random slope.[17]

Study 2 hypotheses are related to SPMM performance via the number of classes supported by the data. Thus, results are primarily analyzed by class rather than by AIC- or BIC- implied solutions (as in Study 1). For the sake of parallelism, Figures 10 and 11 show parameter bias (RB and SB) and RMSE for the fixed slope (Figure 10) and slope variance (Figure 11), when the AIC is used for class enumeration on a replication-by-replication basis.

There is a clear trend for fixed effect estimates to be more efficient as the number of repeated measures increase; however, SB appears worse as the number of repeated measures increase. RB, the more reliable metric in the case of different observation lengths, does not show this trend. There is no clear trend in the effect of observation length on RB for fixed effect estimates, but variance component estimates appear highly biased when only five repeated measures are present. Interestingly, variance component bias is in the upward direction with five repeated measures, whereas variance

[17] For the 60% dropout condition with five observation occasions, it is expected that only 2 repeated measures are observed for the average individual. The random slope contributes to none of the explained variance in the first observation, and only a small amount to the second observation. This hypothesis was explored with individual replications and support was found for this explanation.

components are always downwardly biased when LCM is applied in the presence of random coefficient dependent missingness.

Similar to the LCM, nature of missingness had a larger influence on SPMM-based parameter estimates than the proportion of missingness. The dropout mechanism results in more biased and less efficient fixed parameter estimates; this trend grows larger as the number of observation occasions increases. Interestingly, the reverse trend is apparent for the variance component estimates. This finding warrants a more detailed analysis of the by-class solutions.

Figures 12-17 show average parameter estimates with error bars marking ± 1 empirical SD across the reps that were analyzed for the fixed slope (Figures 12, 14, and 16) and slope variance estimates (Figures 13, 15, and 17), by number of classes. The average number of classes selected by the BIC is marked with a blue triangle for each cell in the figures, and the average number of classes selected by the AIC is marked with a red square. Figures 12 and 13 correspond to conditions with five repeated measures, Figures 14 and 15 correspond to conditions with ten repeated measures, and Figures 16 and 17 corresponds to conditions with twenty repeated measures. Dropout conditions are shown on the top row and erratic conditions are shown on the bottom row. Conditions with 30% missingness are shown on the left and conditions with 60% missingness are shown on the right. Standard error bias in the fixed slope is represented in Table 11 as the ratio of the average standard error estimate to the observed standard deviation of fixed slope estimates.

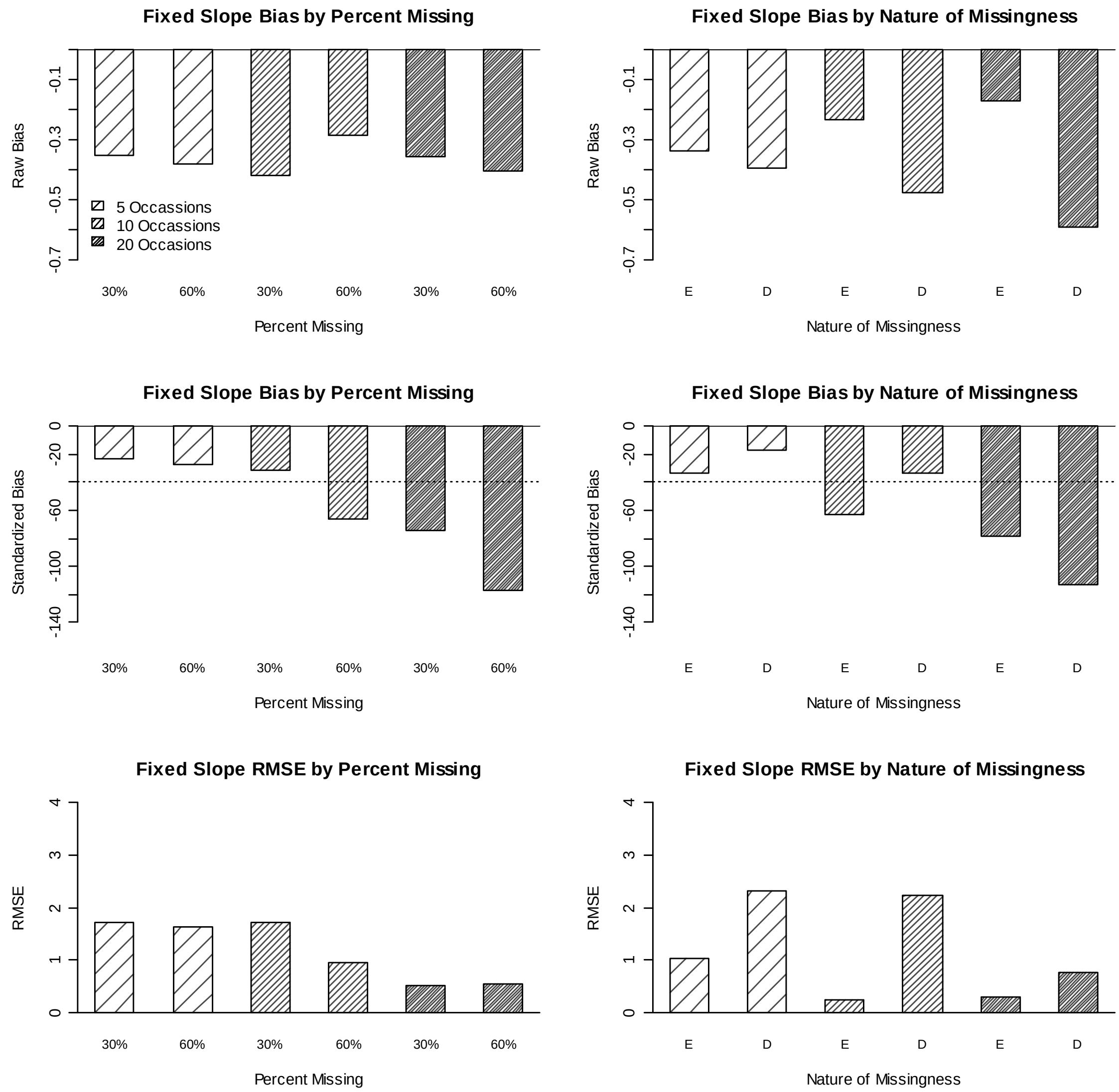


Figure 10. Raw bias (top), standardized bias (middle), and RMSE (bottom) of SPMM-implied fixed effect estimates for the slope by percent of missingness (30% or 60%; left) and nature of missingness (erratic (E) or dropout (D); right) and by number of observation occasions (5, 10, or 20). A horizontal reference line is drawn at SB = -40 to indicate the cutoff for "acceptable" bias.

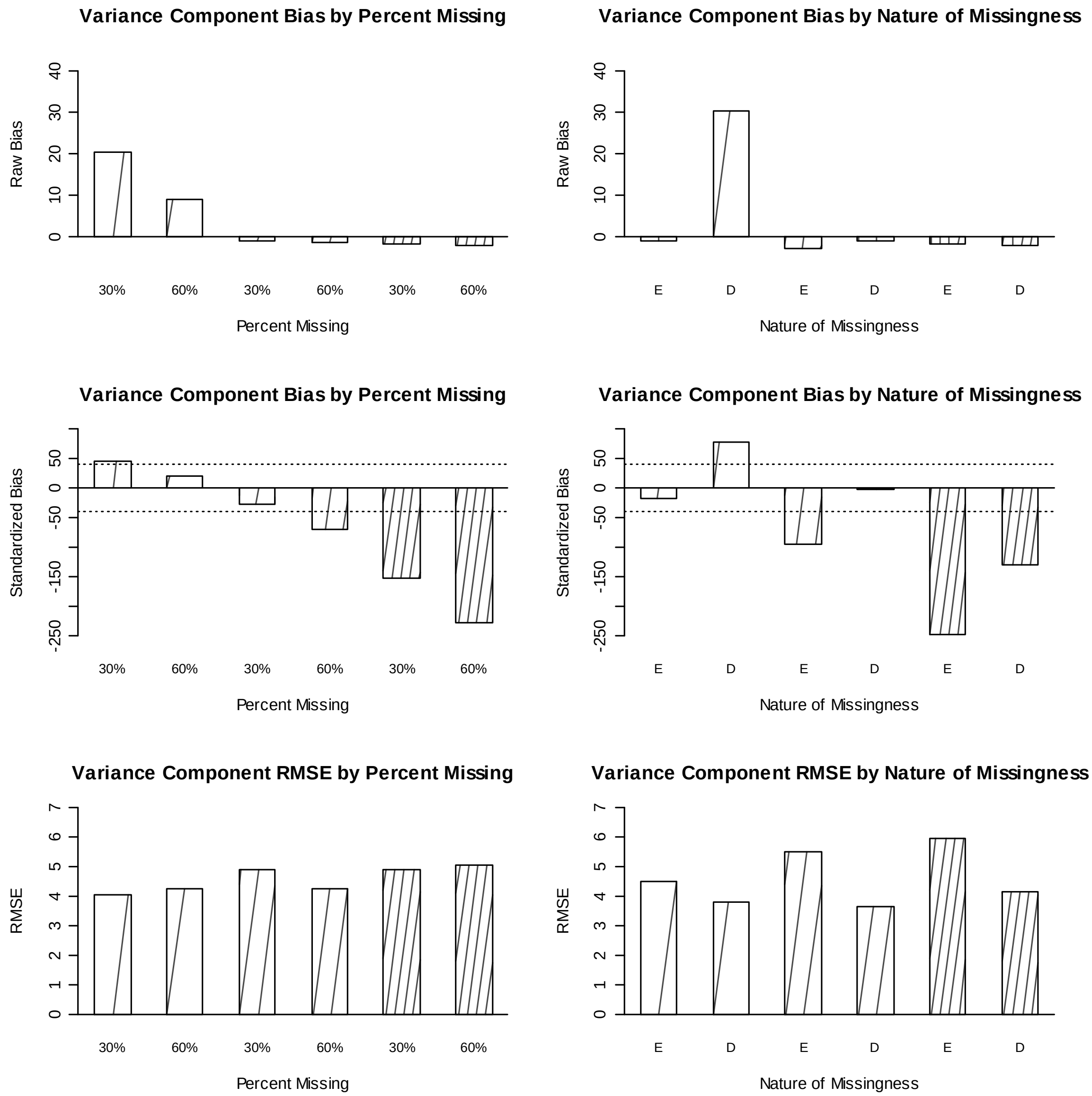


Figure 11. Raw bias (top), standardized bias (middle), and RMSE (bottom) of SPMM-implied variance component estimates for the slope by percent of missingness (30% or 60%; left) and nature of missingness (erratic (E) or dropout (D); right) and by number of observation occasions (5, 10, or 20). A horizontal reference line is drawn at SB = -40 to indicate the cutoff for "acceptable" bias.

Figures 12 and 13 suggest that, practically speaking, only two classes are supported when five repeated measures are present (particularly when a dropout mechanism is operating). Although the AIC and BIC suggest fitting a larger number of classes, a reasonable data analyst would probably not choose to interpret a SPMM with several classes that result in very wide standard error estimates in favor of a solution with fewer classes and smaller standard error estimates, particularly when the fixed effects are stable across solutions. Table 11 provides corroborating evidence that it is not appropriate to extract more than two classes when only five repeated measures are present, particularly when the missingness mechanism is monotone dropout. In the dropout conditions, standard error estimates were over-estimated (over and above the already large empirical standard deviations shown in the figures).[18] Given that the MNAR mechanism imposed in Study 2 was quite severe, these result imply that it may be practical for a data analyst with a small observation length to rely only on two latent classes when conducting a sensitivity analysis for LCM results in the presence of potentially non-randomly missing data.

The second piece of information to glean from Figures 12 and 13 is that fixed effect estimates tend to approach the true parameter value (and rather quickly) as the number of classes increase, but variance component estimates that are initially downwardly biased quickly pass through the true parameter value and become upwardly biased in an unbounded fashion. This explains results displayed in Figure 11; AIC consistently selects too many latent classes, and the BIC selects too many latent classes when 60% of the data are missing due to dropout.

[18] Stable individual differences may not be distinguishable from time-specific noise in these conditions, resulting in upwardly biased standard error estimates.

Table 11**.** *Ratio of Average Estimated Standard Errors to Empirical Standard Deviation for SPMM Fixed Slope Estimates by Study 2 Condition and Number of Classes*

	One	**Two**	**Three**	**Four**	**Five**
5 Observations					
30% Dropout	.96	1.01	7.31	NA	NA
60% Dropout	1.00	.96	8.43	NA	NA
30% Erratic	.96	.98	1.00	1.08	NA
60% Erratic	.94	1.03	.96	.94	NA
10 Observations					
30% Dropout	.97	.93	.96	1.20	1.34
60% Dropout	.96	.98	.86	1.61	.87
30% Erratic	1.00	1.00	1.00	.93	.97
60% Erratic	1.00	1.00	.97	.97	1.00
20 Observations					
30% Dropout	1.04	1.07	1.04	.99	.94
60% Dropout	.96	1.00	.96	.95	1.08
30% Erratic	1.00	.90	.90	.86	.78
60% Erratic	1.00	.90	.86	.86	.86

Note. Five class SPMMs were not estimated for the conditions with five measurement occasions. Sampling distributions for point and standard error estimates were too non-normal to extract meaningful ratio estimates for the four class SPMMs with dropout / five observation occasion conditions

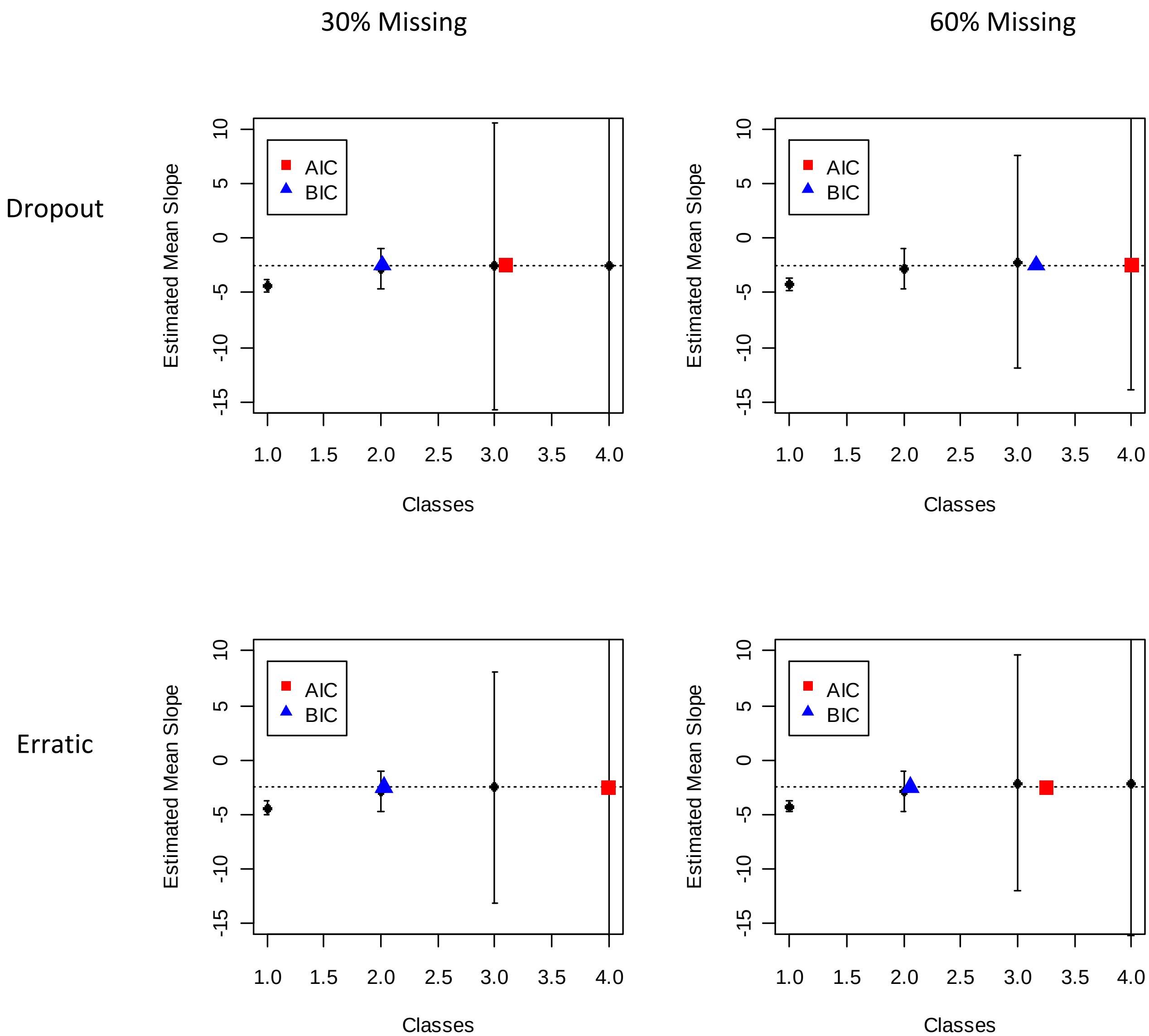


Figure 12. SPMM-implied fixed effect estimates for the slope by study condition (5 repeated measures only). Conditions are: 30% dropout (top left); 60% dropout (top right); 30% erratic missingness (bottom left); 60% erratic missingness (bottom right). Horizontal reference line is drawn at the true parameter. Error bars show ±1 empirical SD.

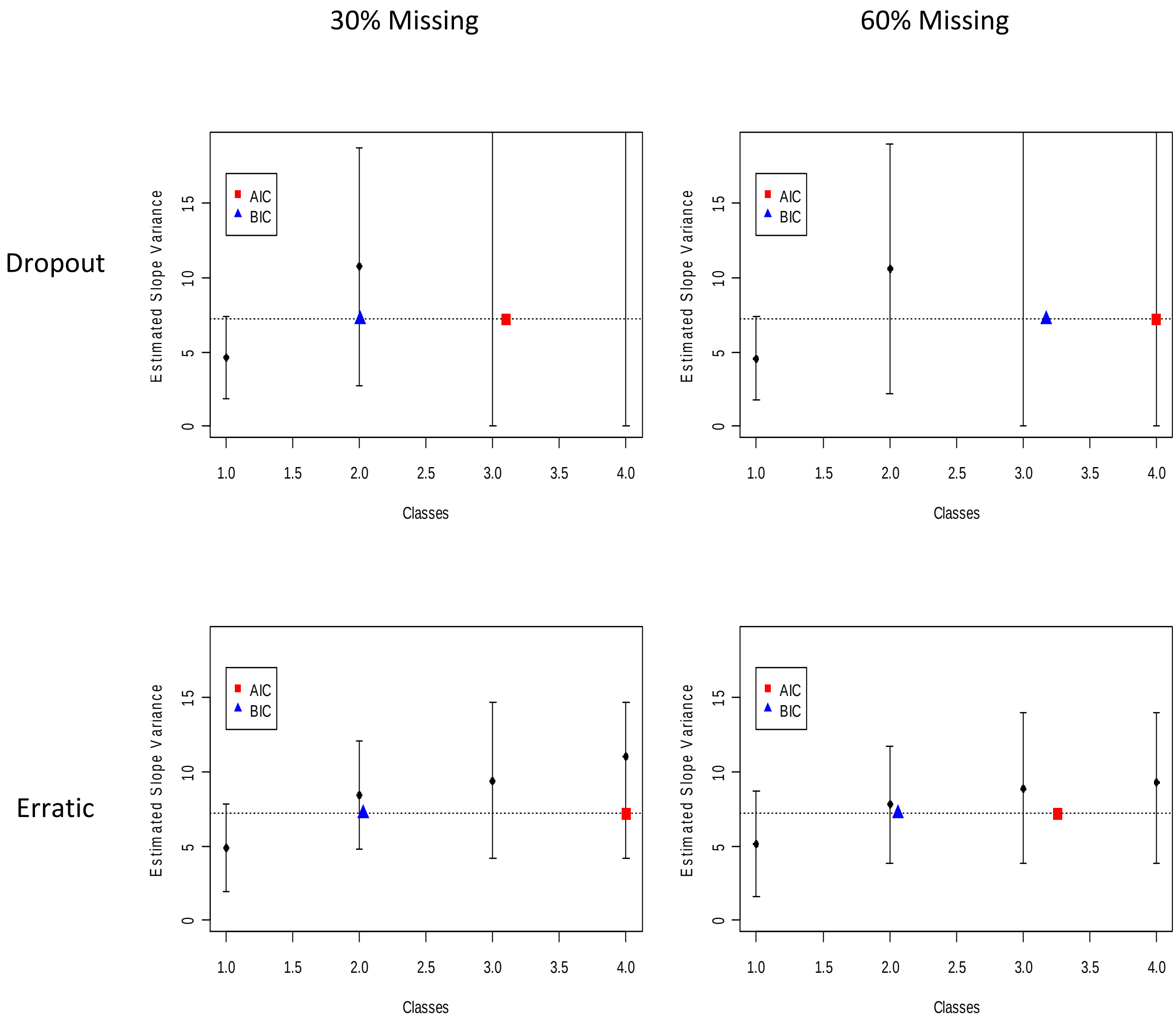


Figure 13. SPMM-implied variance estimates for the slope by study condition (5 repeated measures only). Conditions are: 30% dropout (top left); 60% dropout (top right); 30% erratic missingness (bottom left); 60% erratic missingness (bottom right). Horizontal reference line is drawn at the true parameter. Error bars show ±1 empirical SD.

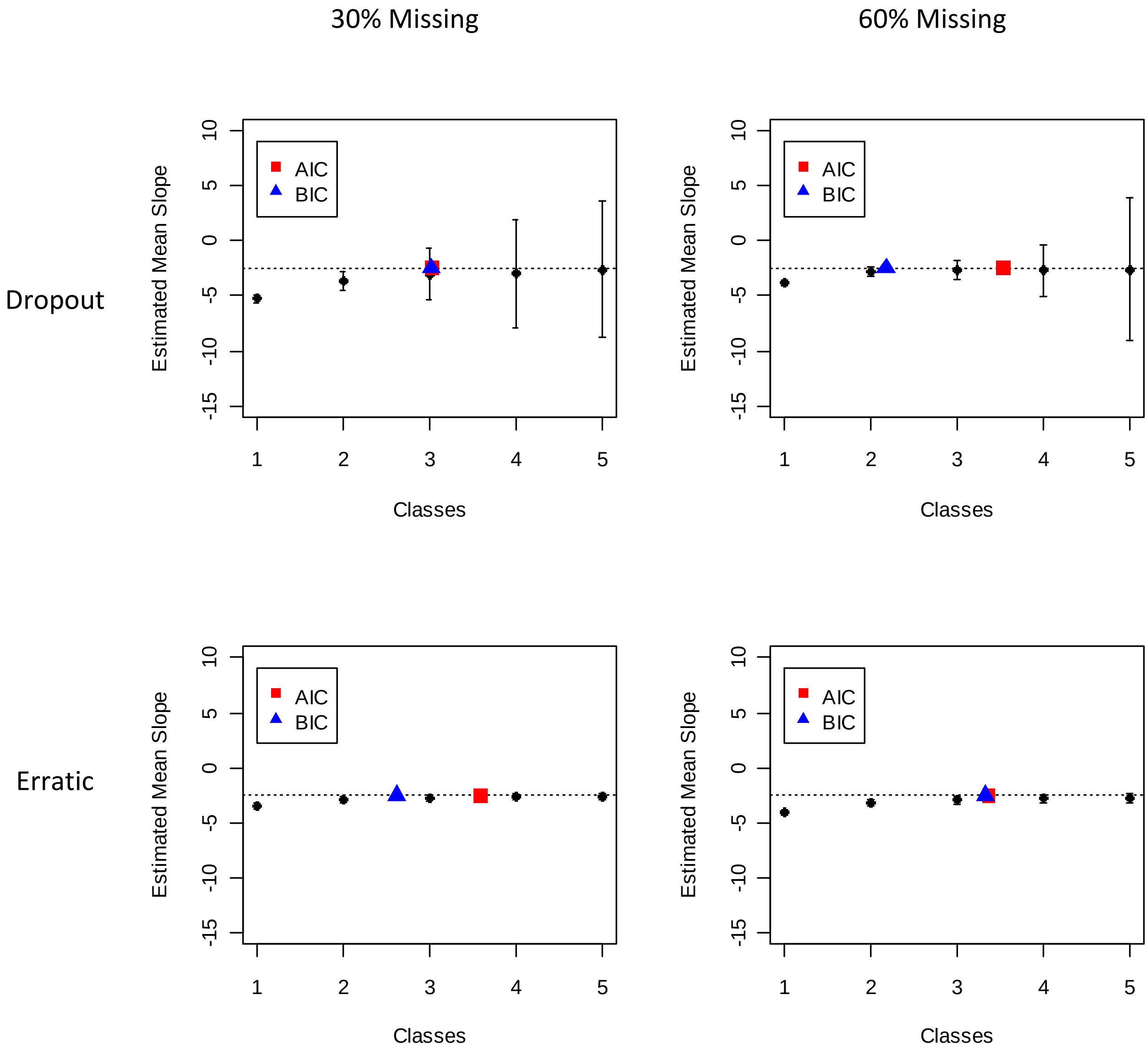


Figure 14. SPMM-implied fixed effect estimates for the slope by study condition (10 repeated measures only). Conditions are: 30% dropout (top left); 60% dropout (top right); 30% erratic missingness (bottom left); 60% erratic missingness (bottom right). Horizontal reference line is drawn at the true parameter. Error bars show ±1 empirical SD.

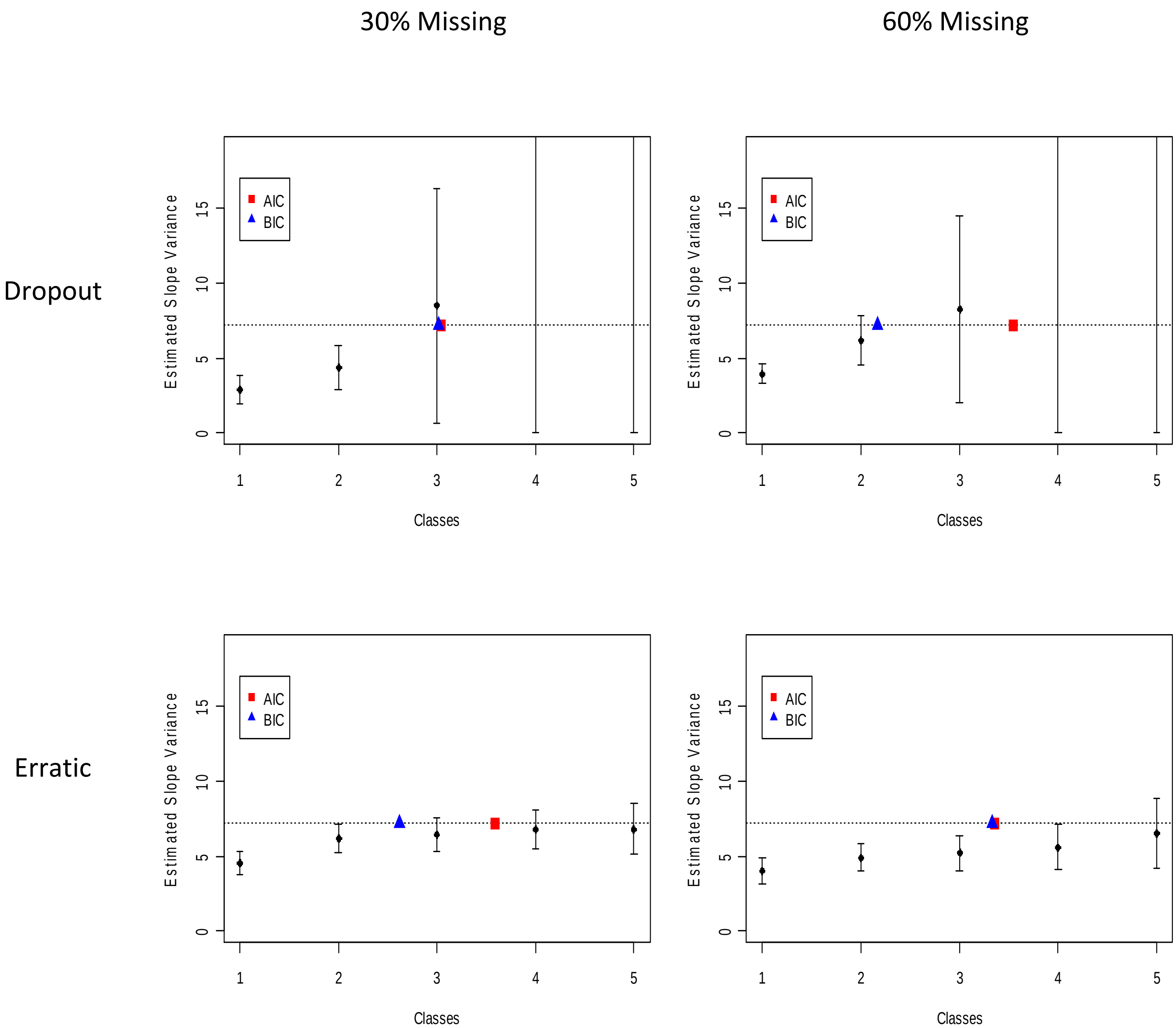


Figure 15. SPMM-implied variance estimates for the slope by study condition (10 repeated measures only). Conditions are: 30% dropout (top left); 60% dropout (top right); 30% erratic missingness (bottom left); 60% erratic missingness (bottom right). Horizontal reference line is drawn at the true parameter. Error bars show ±1 empirical SD.

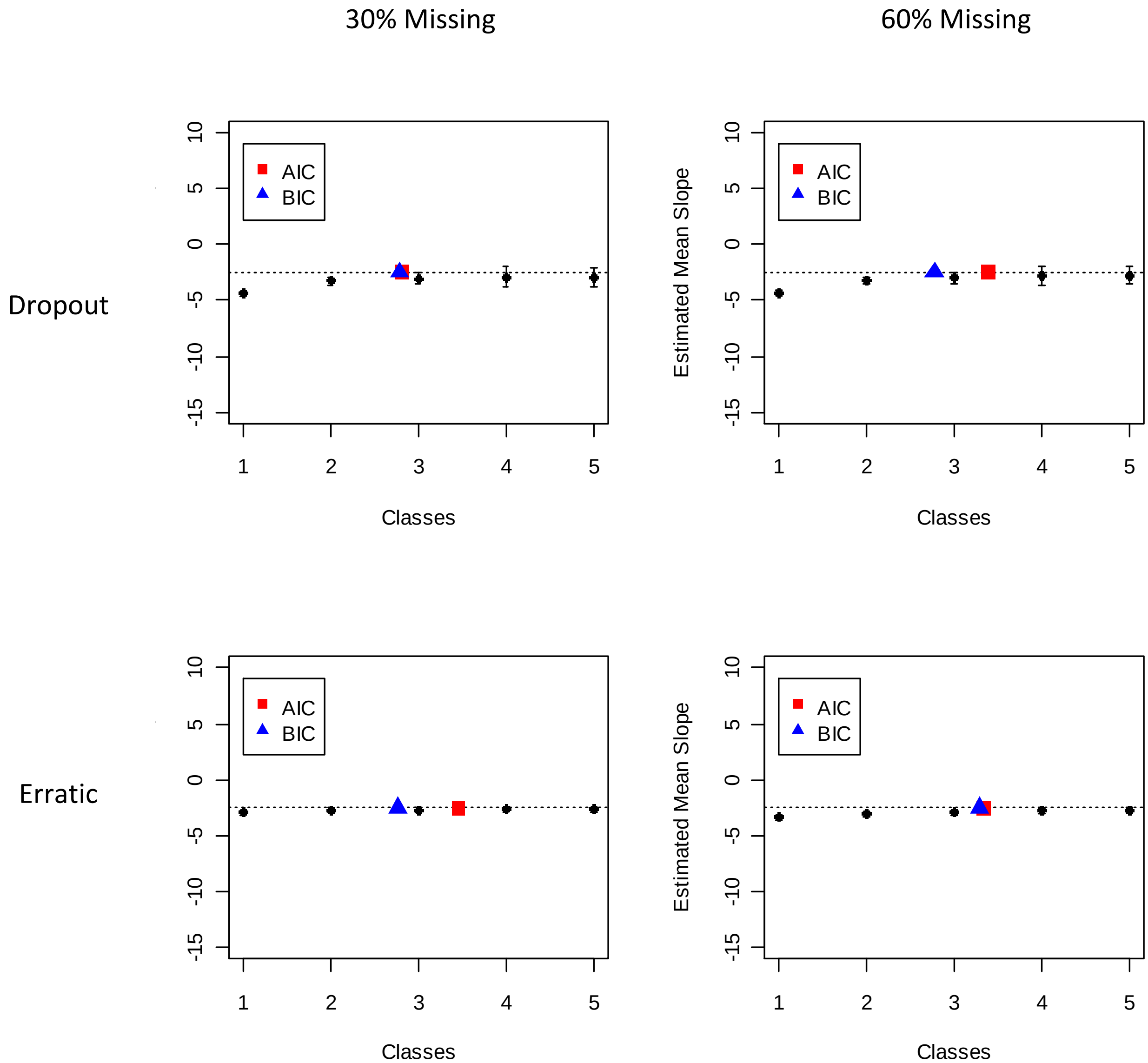


Figure 16. SPMM-implied fixed effect estimates for the slope by study condition (20 repeated measures only). Conditions are: 30% dropout (top left); 60% dropout (top right); 30% erratic missingness (bottom left); 60% erratic missingness (bottom right). Horizontal reference line is drawn at the true parameter. Error bars show ±1 empirical SD.

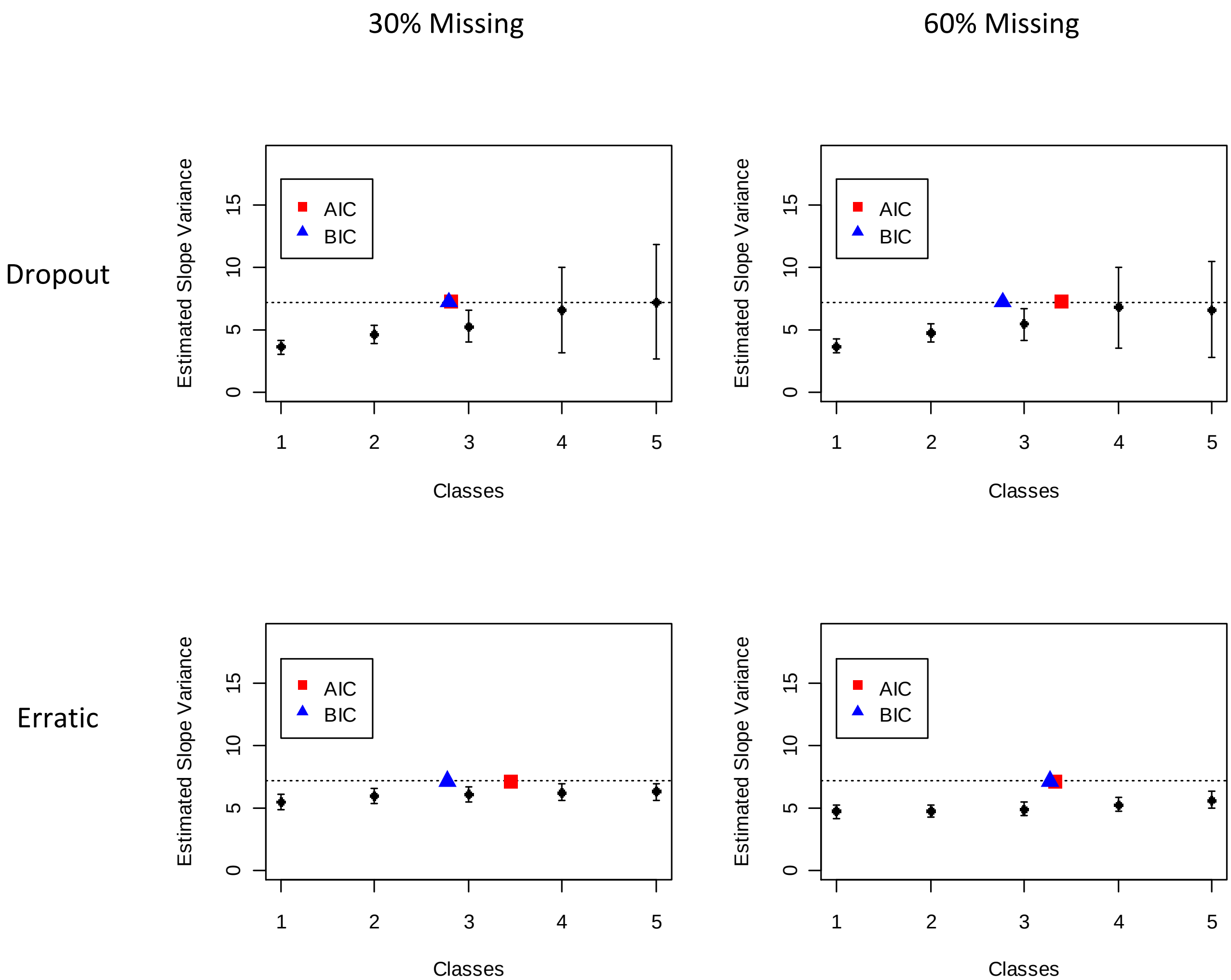


Figure 17. SPMM-implied variance estimates for the slope by study condition (20 repeated measures only). Conditions are: 30% dropout (top left); 60% dropout (top right); 30% erratic missingness (bottom left); 60% erratic missingness (bottom right). Horizontal reference line is drawn at the true parameter. Error bars show ±1 empirical SD.

Moving to Figures 14 and 15 (10 observation occasions), it is immediately apparent that variability of parameter estimates is substantially reduced as five additional observation occasions are added. It is also apparent that SPMM performance is much worse when dropout is present in that the parameter estimates are highly variable across repititions. As was the case with five repeated measures, fixed effects approach the true parameter values as two or three classes are estimated but variance component estimates surpass the true values and increase without bound. Unlike the five repeated measures conditions, standard error estimates were relatively unbiased with ten repeated measures (see Table 11); however, standard error estimates were more biased in the dropout conditions than in the erratic missing data conditions, particularly as the number of classes extracted increased. The proportion of missing data had little influence on fixed effect parameter estimates, but the SPMM had trouble recovering unbiased variance component estimates with a high proportion of missing data. This trend was particularly apparent for the erratic missing data condition.

Finally, moving to Figures 16 and 17 (20 observation occasions), it becomes apparent that observation length indeed has a strong effect on SPMM performance. Specifically, more repeated measures are linked to much less variable parameter estimates. Further, in line with the general finding that fixed effects tend to be more robust than variance components with respect to parameter recovery under the SPMM, it takes fewer repeated measures for the SPMM to obtain good fixed effects estimates than it takes for the SPMM to obtain good variance component estimates.

When 20 observation occasions are collected, fixed effect estimates are quite precise regardless of the nature or proportion of missingness, except that more classes are required to obtain acceptable fixed effect parameter estimates when missingness is due to dropout. Variance component estimates are downwardly biased when there is a high proportion of missing data, particularly with an erratic missingness mechanism.

Standard error estimates are slightly downwardly biased for the erratic missingness conditions, particularly as more classes are estimated. It is interesting that standard errors are more biased in the erratic conditions than in the dropout conditions as the observation length increases; dropout does not appear to be problematic for the SPMM when a large number of repeated measures are present.

Summary and Discussion of Study 2

Although it was not an explicit hypothesis of Study 2, it was implicitly assumed that the SPMM would outperform the LCM under all data conditions that were assessed. This was assumed because the SPMM was better at recovering fixed effect parameter estimates (both in terms of bias and efficiency) in Study 1. However, it was not necessarily expected that the SPMM would recover the variance component of the random slope well based on results from Study 1. Figure 18 provides a direct comparison of the (raw) bias and RMSE of LCM- and SPMM-implied fixed slope (at the minimum BIC because this was determined to be the superior information criterion for most data conditions) estimates by observation length and nature of missingness[19], demonstrating that the SPMM does indeed provide much less biased fixed effect parameter estimates than the LCM across the range of study conditions tests in Study 2. Further, Figure 18 shows that the SPMM is only more efficient than the LCM in the recovery of fixed effect estimates when the missing data are erratically missing, or when many repeated measures

[19] Figures 18 and 19 collapse over the proportion of missing data since this manipulation had little influence on parameter estimates.

are present and missingness is due to monotone dropout. In other words, the SPMM may be more accurate than the LCM in the recovery of fixed effects *on average*, but SPMM-implied results are too variable to be of much use when dropout is present and when there are not many repeated measures.

Figure 19, which shows a comparison of LCM- and SPMM-implied variance components, tells a different story. LCM-based variance component estimates are less biased and less variable than SPMM estimates. On the other hand, results presented in these figures are based on results obtained using the BIC for class enumeration. Results presented in Study 2 suggest that the BIC (and the AIC) *overestimate* the number of classes necessary to achieve approximate conditional independence between the missing data and the repeated measures; this over-extraction has serious consequences for inducing excessive variability in parameter estimates and for inducing positive bias in variance component estimates.[20]

Study 2 presented compelling evidence to suggest that the AIC and BIC are not optimal criteria for enumerating classes in an SPMM context. Performance of these indices depends on the number of repeated measures present in a sample such that too many classes are selected when the number of repeated measures is small and too few classes are selected when the number of repeated measures is large. The equations for calculating AIC and BIC (shown in Equation (2.8)) do not explicitly account for the number of level 1 units. The BIC penalizes for sample size at level 2; in this case, however, it appears that a large level 1 sample size should be 'rewarded' (rather than

[20] If parameter bias and efficiency are used as optimal criteria, then study results suggest that it *may* be pragmatic to utilize fixed effect estimates from the SPMM and variance component estimates from the LCM. However, this practice is not justifiable from an analytical standpoint. Future research should evaluate the theoretical rationale and the consequences of such an approach.

penalized). Other fit indices exist for comparing solutions (e.g., the Lo-Mendell-Rubin likelihood ratio test, Lo, Mendell, & Rubin, 2001; the consistent AIC, Bozdogan, 1987; the skewness and kurtosis tests; Muthėn, 2003; sample size adjusted BIC; Sclove, 1987); however, it is doubtful whether any of these fit indices would be an improvement over the AIC and BIC because level 1 sample size is not explicitly considered by any of these criteria.

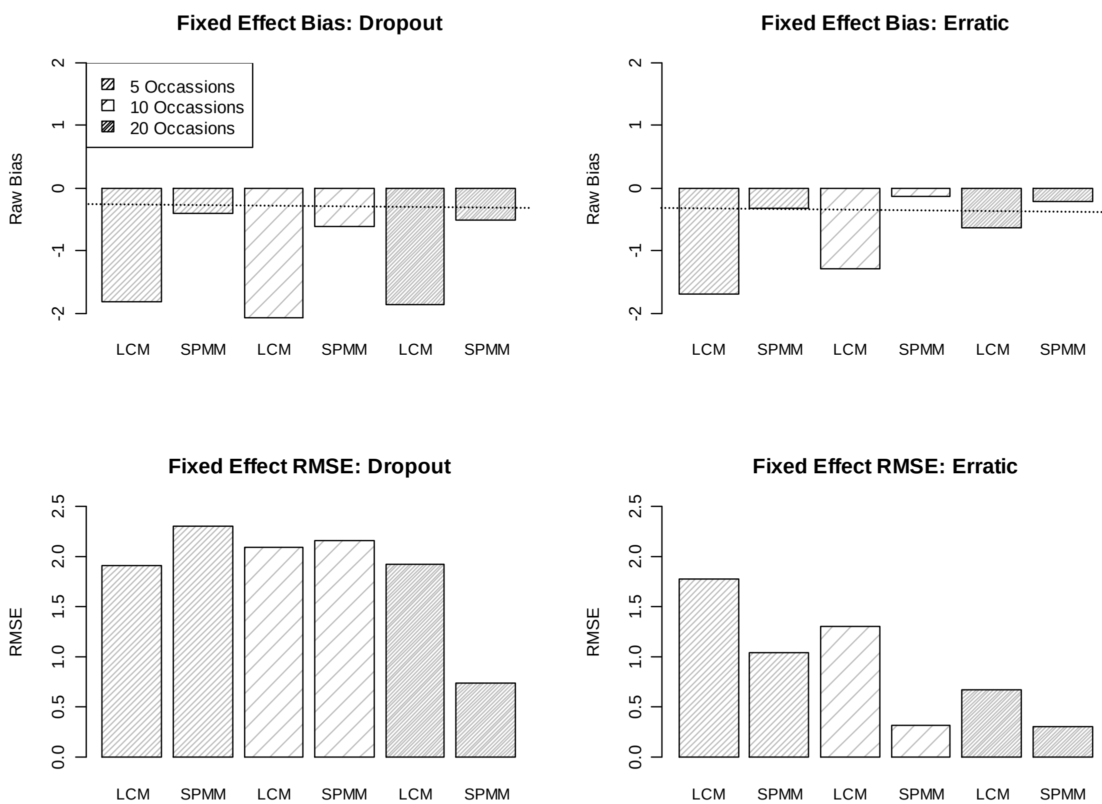


*Figure 18***.** Comparison of bias (top) and variability (bottom) in LCM- and SPMM-implied fixed effects, by nature of missingness (i.e., dropout (left) or erratic (right)) and observation length. A dotted line is draw to represent -10% raw bias.

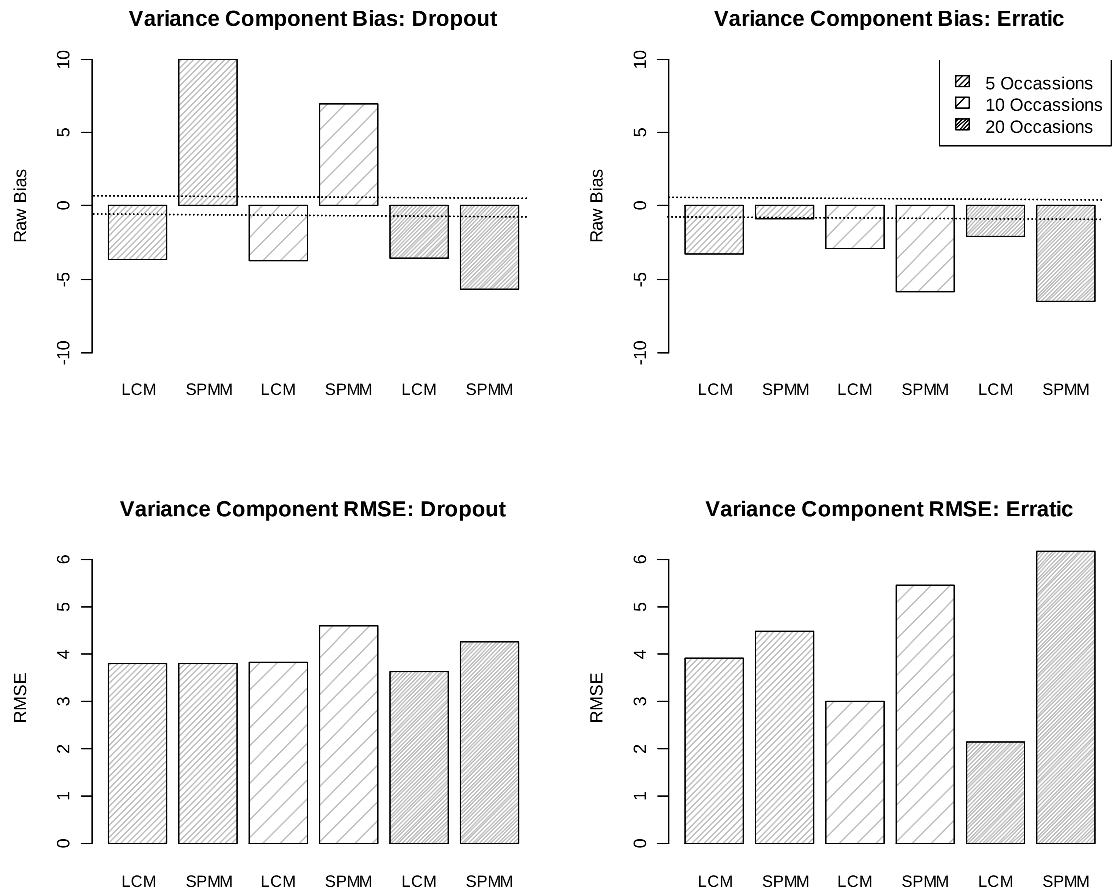


Figure 19. Comparison of bias (top) and variability (bottom) in LCM- and SPMM-implied variance components, by nature of missingness (i.e., dropout (left) or erratic (right) and observation length. A dotted line is draw to represent ±10% raw bias.

Lin et al. (2004) presented a method for assessing the conditional independence assumption in a latent mixture model that has good potential for use as an alternative method for class enumeration in the SPMM context. Lin and colleagues suggested generating *M* {*M* = 1,….m}sets of randomly generated multinomial vectors indicating latent class membership, ***D***, where *K* is fixed, based on individuals' posterior probabilities of class membership (which are automatically generated when the EM algorithm is used to estimate the model). The *M* data sets with new ***D*** indicators are then stacked to form a single data set, assigning a weight of 1/*M* to each case. Each individual in the original sample then has *M* imputed copies of an 'observed' class membership variable. If the conditional independence assumption is met, the missing data indicator(s), $\boldsymbol{R}$, should be uncorrelated with the repeated measures, $\boldsymbol{Y}$, conditional on imputed class membership, $\boldsymbol{D}$. This can be evaluated by estimating a growth model (e.g., an LCM) with $\boldsymbol{D}$ as an observed predictor of the growth factor means $\boldsymbol{\alpha}$ (similar to a multiple groups model). The missing data indicator(s) $\boldsymbol{R}$ are also included as predictors of growth, but their estimated effect should be zero if conditional independence exists. This approach has the benefit of being directly relevant to the goal of class enumeration in SPMM; however, its performance has not been evaluated with simulation methodology. Since Study 2 showed that more classes are necessary to achieve adequate bias reduction as the number of observations increase (indicating that conditional independence is achieved more quickly with fewer repeated measures), it is likely that Lin et al.'s approach would work well when used in conjunction with traditional measures like the BIC.

Results from Study 2 suggest that LCM and SPMM performance is highly dependent upon data conditions, even when the missing data mechanism is RC-MNAR-M with severity held constant. LCM is never protected from the negative effects of non-randomly missing data, but data conditions leading to less growth factor determinacy (i.e., dropout as opposed to erratic missingness and a small number of observation occasions) make parameter estimates even less trustworthy. Unfortunately, these are the same conditions that are troublesome for the SPMM. In light of these results, it is necessary to outline some practical guidelines for longitudinal researchers who suspect that their missing data might be missing due to a random coefficient dependent mechanism.

The first piece of advice is to maximize growth factor determinacy. This means establishing reliable measures, collecting as many repeated measures as possible, and trying to collect information from individuals throughout the range of the study (i.e., avoid monotone dropout, for instance by investing resources in locating persons who might otherwise be lost to follow up). Particularly when fewer repeated measures are planned, it is important to avoid having individuals who drop out after the first few occasions.[21] A researcher with limited resources to spend on data collection might consider a planned missingness approach; Study 2 results showed that estimates do not suffer when random missingness is added, particularly if random missingness enables less non-random missingness to be present.

Once data have been collected, if random coefficient dependent missingness is suspected, it is always a good idea to conduct a sensitivity analysis of LCM results using a SPMM. If fixed effects are similar, then LCM results are probably trustworthy unless OD-MNAR is a possibility. If fixed effects differ, and if more than five repeated measures were collected or if only five repeated measures are present but missing data are erratically spaced, then SPMM-implied fixed effect estimates are probably more trustworthy than LCM estimates. If fixed effects differ but dropout is present with few repeated measures, then it may be wise to conduct further sensitivity analyses using other

[21] Although dropout did not plague parameter estimates as much when more observations were collected in the simulation study, researchers should be careful about extrapolating beyond the range of the available data during 'real' data analysis when the functional form cannot be assumed to remain constant for an indefinite period of time.

methods for handling non-randomly missing data (e.g., traditional PMMs or traditional shared parameter models with a variety of parametric assumptions imposed).

Fixed effects are more robust to the number of classes extracted in SPMMs than variance components. That is, fixed effects estimates are generally unbiased if two or more classes are extracted in a SPMM (although more classes are recommended if many repeated measures are observed). Particularly if erratic missingness is present, or if there are many repeated measures, variability of the fixed effects estimates is not too large. On the other hand, variance components are highly sensitive to the number of classes that are extracted in an SPMM. If variance components are of interest, it may be wise to rely on as few classes as possible, even fewer than are suggested by the information criteria.

The finding that the AIC, and particularly the BIC, tend to over-extract classes in the sense that they sacrifice efficiency far too much, in some cases, was surprising. It was expected that the AIC would be the preferred metric for class enumeration because of its tendency to extract more classes than the BIC. Instead, even the BIC was sometimes too lenient in class extraction. Perhaps this finding would be different if a larger sample size were assessed (N = 300 is rather small, particularly when 30% - 60% of observations are missing), or if the number of observations per person were considered by the information criteria. In practice, it may be difficult for a researcher to determine whether fixed effect parameters that change as the number of classes increase are changing because additional latent classes are reducing bias by accounting for the missing data process, or if they are changing because of additional variability that is being introduced by extra model parameters. This problem may be aggravated with a growth model that is more complicated than the model tested here. A researcher in this situation should monitor standard error estimates across SPMM solutions; if a solution with fewer classes has much smaller standard errors than a solution with more classes, results from Study 2 suggest that the model with fewer classes may be a better choice.

It is interesting to consider what results might have looked like if all of the missing data mechanism that were considered in Study 1 were also assessed in Study 2. In all likelihood, results would have been quite similar across the range of mechanisms. This is because the issue of growth factor determinacy is somewhat orthogonal to the conditions studied in Study 1. Specifically, growth factor determinacy should help the LCM generate better parameter estimates regardless of the missing data mechanism and the SPMM should always perform better with more information.

Chapter 4

STUDY 3: A CASE STUDY OF SPMM: INFERRING CHANGE OVER TIME IN A SAMPLE OF PSYCHOTHERAPY PATIENTS

Chapters 2 and 3 considered the performance of the SPMM with artificial data. In this chapter, the SPMM is applied to a real-data example to show how the model might be usefully applied in practice.

Data from this chapter are the same as those used by Baldwin et al. (2009), who analyzed data from a longitudinal study of psychotherapy treatment in a naturalistic setting. In the Baldwin et al. manuscript, the authors critiqued the long-standing tradition of dose-effect models of change in psychotherapy, a model which aggregates across individuals (assuming no individual differences in rates of change) to test the incremental value added of additional psychotherapy sessions. Baldwin and colleagues were concerned that the dose-effect models assume that response to therapy is equal across individuals, regardless of their dose. Instead, the authors thought that individuals who receive a small dose of therapy might experience the largest effect per dose, and individuals receiving a large dose of therapy might receive the smallest effect per dose. By including a covariate quantifying the total number of sessions attended by each participant in the growth model, and by allowing that term to interact with an individual's rate of change, Baldwin et al. found support for their hypothesis. Ultimately, Baldwin and his colleagues concluded that it is not appropriate to establish a fixed standard of treatment for individuals, given individual differences in dose-effect response.

It is possible to view Baldwin et al.'s (2009) problem as one of potentially random coefficient dependent missing data. If one wishes to draw inferences about average individual trajectories and about variability in individual trajectories *during* psychotherapy treatment, which was the purpose of Baldwin et al.'s analysis, then data are 'missing' from patients after they leave psychotherapy treatment, and the data may be missing due to a non-random, random coefficient dependent mechanism. Viewed in this context, it becomes apparent that Baldwin et al.'s approach of including number of sessions attended as a measured indicator of 'missingness' is akin to a traditional pattern mixture modeling approach to handling missing data. This approach was criticized by Demirtas and Schafer (2003) because the explicit inclusion of an indicator of missingness limits growth model generalizability. In other words, validity of inferences regarding individual- and population-level growth in Baldwin et al.'s model is conditional on the number of sessions attended in the precise way that the variables were entered into the model (i.e., as a main effect and as an interaction with time (measured in session units). Including number of sessions attended as a fixed covariate of growth implies homogeneity of growth trajectories for all individuals who attended the same number of sessions, and it implies heterogeneity of growth for individuals attending a different number of sessions (Hogan & Laird, 1997).

The SPMM might be a good alternative strategy for analyzing Baldwin et al.'s (2009) psychotherapy data while enabling the *un*conditional interpretation of growth coefficients in the population (i.e., marginal over individual differences in total number of sessions attended). This is beneficial for two reasons. First, the explicit model specified by Baldwin and colleagues may be incorrect; there might be a nonlinear main effect of total sessions attended, for instance. Second, it is more useful to have a model predicting change over time that does not require a priori knowledge of total sessions attended by a patient. Consider a scenario in which a patient is consulting with a therapist to determine whether he or she might be a good candidate for receiving psychotherapy treatment, and the patient enquires about how quickly they can expect to experience clinically significant improvement. Under the conditional model, the therapist would not be able to answer the question easily. Under the SPMM, the therapist would be able to tell the patient an average rate of change over time given the prospective patient's initial level of psychological symptoms and other (observable) background factors. Further, the therapist could give an estimate of the average variability in rates of change over time.[22,23]

This chapter walks through a data analytic strategy for the psychotherapy data that were previously analyzed by Baldwin et al. (2009). Data analytic decision points are informed by results from Chapters 2 and 3. The goal of this chapter was to conduct a sensitivity analyses about the inferences drawn regarding the expected shape of change in psychological symptoms over the course of psychotherapy, and about the population variation around the average trajectories.

---

[22] The same information is available from an LCM, but LCM-based inferences would only be accurate if there were no individual differences in dose-effect response, which Baldwin et al. (2009) deemed to be false.

[23] Alternatively, a survival model might be used to predict time to dropout. Lin et al. (2004) presented a joint model for growth over time and survival time; however, this can be viewed as a type of selection model that requires an explicit specification of the dropout mechanism. Future research should assess Lin et al.'s model under misspecification using a simulation design.

The first step in the data analytic plan was to follow a standard procedure for analyzing change over time by analyzing the psychotherapy data using a LCM. Then a SPMM was considered for the same data to evaluate the sensitivity of LCM results to a suspected random coefficient dependent missingness mechanism. Results from Chapter 2 suggest that the SPMM will be useful for identifying and correcting for nonignorable missingness resulting from a RC-MNAR mechanism, such as the one suspected here. Chapter 2 suggests that fixed effect estimates should be quite accurate under the SPMM, but that variance component estimates are likely to be biased to some degree, particularly if multiple nonignorable mechanisms are acting simultaneously (e.g., if both very fast and very slow recovery are related to dropout from therapy).[24] LCM-based estimates of the slope mean and variance component are expected to be biased to the extent that random slope dependent missingness is present. Chapter 3 results suggest that the monotone nature of dropout in the psychotherapy data may lead to excessive variability in parameter estimates, particularly the variance components. However, the average individual in the sample has six or seven observed repeated measures, indicating that at least the fixed effect estimates should be trustworthy under the SPMM.

Methods

Participants and Procedure

Participants were drawn from an archival dataset of therapy outcomes that is maintained by a large university counseling center. Participants in this study were completing their first round of individual psychotherapy, meaning that the longest interval between therapy sessions had not exceeded 90 days. Patients who attended at

[24] Patients who attended fewer than three sessions were not included in the sample, a decision which likely removed the majority of study participants who would not benefit from therapy.

least three, but no more than 27 sessions were included in the sample (most clients attended fewer than 27 sessions and those attending only two sessions did not provide enough information to calculate growth trajectories).[25] Patients discontinued therapy at their (and their therapists') discretion. Only data from the first 14 sessions are analyzed because the majority of patients had dropped out of therapy by this point.

Of the 4,676 patients analyzed in the sample, the average length in treatment was 6.46 sessions (SD = 4.15). The majority of patients had adjustment disorders (37.96%), mood disorders (24.59%), or anxiety disorders (12.13%). Most patients were single (65.06%), White (89.07%), and female (62.32%). Ages ranged from 17-60, with a mean age of 22.28 (SD = 3.70). There were 204 therapists treating the sample of patients, but many patients were seen by multiple therapists over the course of treatment.

Psychotherapy Outcome Measure

The Outcome Questionnaire-45 (OQ-45), a 45-item self-report sum score measure of overall psychological functioning (Lambert, Morton, Hatfield, et al., 2004), was used to assess patients' symptom trajectories over time. The measure assesses three domains: subjective discomfort (e.g., "I feel blue"), inter-personal relationships (e.g., "I feel lonely"), and social role performance (e.g., "I have too many disagreements at work/school"). Possible scores range from 0 (high functioning) to 180 (low functioning), and they ranged from 0 to 166 in this sample. The OQ-45 has been shown to have high internal consistency, test-retest reliability, and concurrent validity (Snell, Mallinckrodt, Hill, & Lambert, 2001; Lambert et al., 2004; Baldwin et al., 2009;). Participants completed the OQ-45 at initial intake and prior to each therapy session.

[25] Excluding patients who only attended one or two sessions may limit generalizability of the study results. Early drop-outs may differ from people who remain in therapy for more than two sessions.

Data Analysis Plan

Baldwin et al. (2009) estimated a cubic growth model because the rate of growth was not linear (i.e., the rate of improvement slows over time).[26] In this analysis, a variety of unconditional growth models were assessed using an LCM before settling on a log-linear model for time. The following LCM was estimated as a second step in the modeling process:

$$OQ45_{ti} = \eta_{0i} + \eta_{1i} * \ln(Session) + \varepsilon_{ti}$$

$$\begin{aligned} \eta_{0i} &= \alpha_0 + \lambda_{01} * Anx\_Dx_i + \lambda_{02} * Mood\_Dx_i + \lambda_{03} * Other\_Dx_i \\ &+ \lambda_{04} * NotSingle_i + \lambda_{05} * Minority_i + \lambda_{06} * Age_i + \lambda_{07} * Male_i + \zeta_{0i} \end{aligned} \tag{4.1}$$

$$\begin{aligned} \eta_{1i} &= \alpha_1 + \lambda_{11} * Anx\_Dx_i + \lambda_{12} * Mood\_Dx_i + \lambda_{13} * Other\_Dx_i \\ &+ \lambda_{14} * NotSingle_i + \lambda_{15} * Minority_i + \lambda_{16} * Age_i + \lambda_{17} * Male_i + \zeta_{1i} \end{aligned}$$

with the following random effect distributions:

$$\begin{gathered} \begin{bmatrix} \zeta_{0i} \\ \zeta_{1i} \end{bmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \psi_{00} & \psi_{01} \\ \psi_{10} & \psi_{11} \end{pmatrix}\right] \\ . \\ \varepsilon_{ti} \sim N(0, \sigma^2) \end{gathered} \tag{4.2}$$

In this analysis and in Balwin et al.'s (2009) analysis, time was measured in session units rather than in chronological time. Session attendance was not evenly spaced across individuals, but the sample was restricted to individuals who did not allow a long lapse of time (90 days or more) in between psychotherapy sessions. Sessions were the preferred unit of time for this analysis because it was the most relevant to the substantive

[26] This is a well-established trend in literature evaluating response to psychotherapy treatment.

question of interest: what is the typical rate of response to the receipt of psychotherapy, and how much individual variation exists in response to therapy?

Anxiety disorder, mood disorder, and other diagnoses were included as predictors of the random intercept and of the random slope, with "adjustment disorder" as the reference group. A relationship status of "married" or "cohabitating," identification as an ethnic minority, participant age (grand mean centered), and being male were included as time invariant covariates of growth as well, with single people, Whites, and females as referents. These predictors were all included in Baldwin et al.'s (2009) analyses.

The next step of the analyses involved fitting a series of SPMMs to the data, increasing the number of latent classes as necessary. This involved fitting one- through three-class SPMMs to the data. A decision to stop adding classes was made after small class proportions appeared in the four class model (π_k = .05), and because aggregate parameter estimates did not change substantially as more classes were added. A single summary indicator was used to indicate the 'dropout' occasion-- the log of the total number of sessions attended. Total number of sessions attended ranged from 3 – 27 for the sample and was heavily skewed right. Log number of sessions was distributed with Mean = 2.05, Median = 2.08, SD = .60, Skew = .07, and Kurtosis = -.74.

Mplus version 6 was used to estimate the models, and the Model Constraints command was used to calculate the population-average intercept, slope, and variance component parameter estimates so that standard errors would be output by the program. Delta method standard error estimates were obtained because bootstrapped standard errors are not yet available for this type of model in Mplus software. Results from

Chapter 3 suggest that standard error estimates should be trustworthy, particularly if a conservative approach is taken to class enumeration.

RESULTS

LCM Results

The LCM fit relatively well ($\chi^2(197) = 2029.558,\ p < .001$; CFI = .941; TLI = .943; RMSEA = .045). Parameter estimates are presented in Table 12.

Table 12. *LCM-Implied Parameter Estimates*

	Estimate	SE
Fixed Effects		
Intercept	67.83***	(.63)
Slope	-6.78***	(.35)
Intercept Regression Parameters		
Anxiety	6.97***	(1.02)
Mood	17.85***	(.80)
Other	1.29	(.80)
Male	-3.95***	(.66)
Not Single	-2.94***	(.66)
Minority	4.45***	(1.00)
Age	.12	(.09)
Slope Regression Parameters		
Anxiety	-.54	(.56)
Mood	-3.35***	(.44)
Other	.69	(.44)
Male	.59	(.36)
Not Single	-1.29***	(.36)
Minority	-1.09*	(.55)
Age	.11*	(.05)
Variance Components		
Intercept Variance	376.69***	(9.29)
Slope Variance	73.37***	(2.73)
Covariance	-45.26***	(3.84)
Residual Variance	102.80***	(1.03)

Note. * p <.05; *** p<.001

Results from the LCM analysis indicate that psychological symptoms decrease steadily throughout the 14 sessions that were analyzed, but that the rate of change declines as sessions increase. The LCM-implied trajectories for single, White females (the majority of the sample), of average age, with anxiety disorders, mood disorders, adjustment disorders, and all other disorders are illustrated in Figure 20.

Further, results suggest that the average patient entering therapy with an anxiety disorder is more severe than a patient entering with an adjustment disorder, and a patient entering therapy with a mood disorder is dramatically more severe, initially, than a patient presenting with an adjustment disorder. On average, males, Whites, and people who are married or cohabiting begin therapy with less severe psychological symptoms than women, ethnic minorities, or single people.

People with adjustment disorders, anxiety, or other disorders (except for mood) improve at about the same rate on average. People with mood disorders tend to improve more quickly than the other groups. People who are married or cohabitating and ethnic minorities tend to improve more quickly with therapy than people who are single or White.

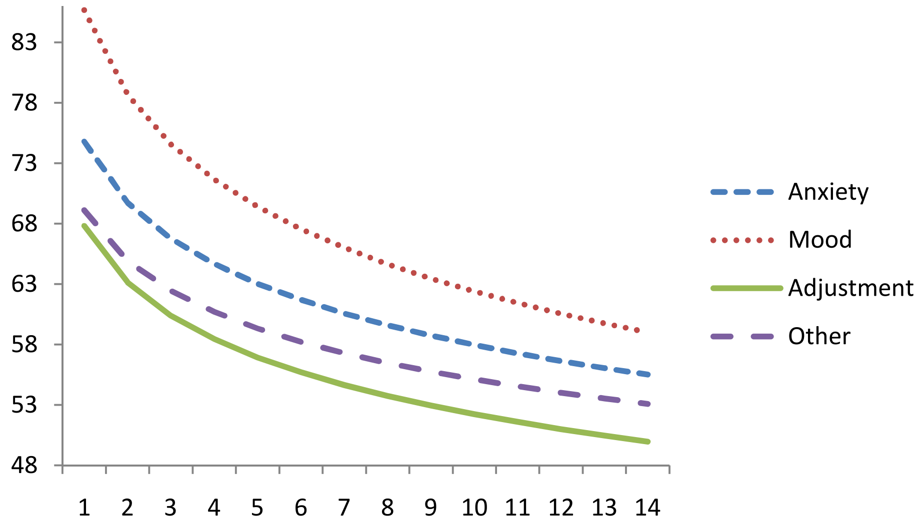


Figure 20. LCM-Implied Average Trajectories for Psychological Symptoms for White, Single, Female Psychotherapy Patients

The LCM results suggest that there is substantial individual heterogeneity around baseline psychological functioning, even after accounting for the observed individual differences in diagnosis and demographic characteristics. There is also substantial heterogeneity with respect to rate of change over time, and individuals who enter psychotherapy with more severe psychological symptoms are likely to improve more quickly than people who enter therapy with less severe symptoms. There is also a substantial amount of unexplained, time-specific residual variance.

Sensitivity Analysis with SPMM

To the extent that individuals who improved more quickly over time (after accounting for observed covariates, including diagnosis, age, gender, ethnicity, relationship status, and baseline OQ-45 scores) were also the patients who left therapy early (as suggested by Baldwin et al., 2009, who were the first to analyze this psychotherapy data), results from the LCM will be inaccurate. Based on results from

Chapters 2 and 3 of this manuscript, it is expected that the fixed slope estimate generated by the LCM is biased to some degree, as well as the estimated variance of the slope and covariance between the intercept and slope.

As shown in Figure 21, AIC and BIC continued to improve as more classes were added, as often happens with large samples. However, results from Study 2 suggest that the AIC and BIC may sometimes overestimate the number of classes necessary to approximate the missingness mechanism. Further, results suggested that it may be more problematic to utilize too many classes than too few in terms of RMSE / efficiency loss. Finally, it is clear from Table 13 that parameter estimates are remarkably similar for the two- and three- class solutions. As expected, inference regarding predictor effects is not changed across models, even when compared with LCM-based inference. This concurs with findings from Study 1.

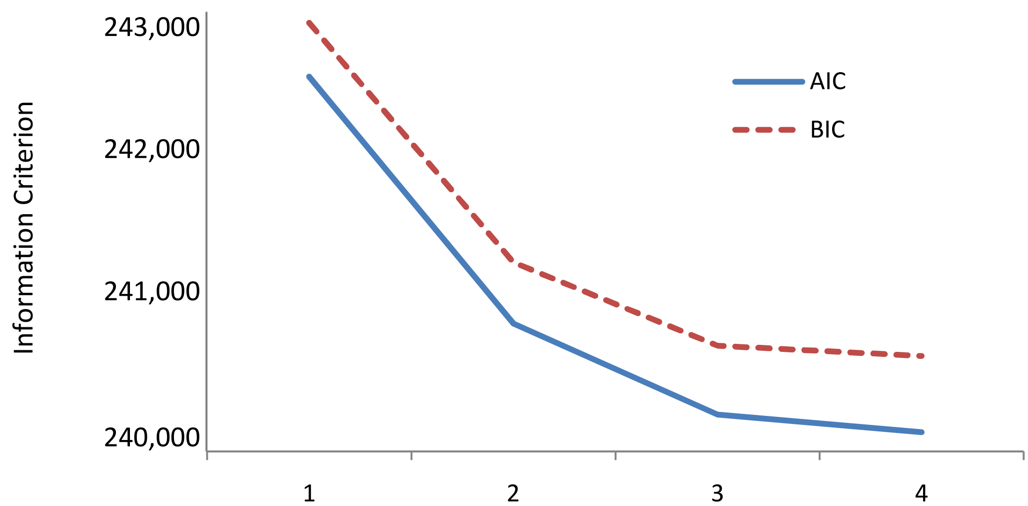


Figure 21. AIC and BIC values as a function of the number of latent classes. This plot suggests dramatic improvement in fit when moving from a one- to two-class solution, and slight improvement in fit when moving from a two- to three-class solution.

Because the population average intercept term remains relatively constant across models while the average slope changes, these results are consistent with a mild random slope-dependent MNAR process. For comparison, one- through three-class SPMM-implied trajectories for single, White females of average age are compared in Figure 22.

Table 13. *One- through Three-Class SPMM-Implied Psychotherapy Trajectory Estimates*

	1 Class SPMM		2 Class SPMM		3 Class SPMM	
	Estimate	SE	Estimate	SE	Estimate	SE
Fixed Effects						
Intercept	67.83***	.63	68.12***	.63	68.16***	.63
Slope	-6.78***	.35	-7.24***	.35	-7.34***	.35
Intercept Regression Parameters						
Anxiety	6.97***	1.04	6.93***	1.04	6.93***	1.04
Mood	17.85***	.78	17.73***	.78	17.68***	.78
Other	1.29	.81	1.21	.81	1.22	.81
Male	-3.95***	.66	-3.92***	.66	-3.90***	.66
Not Single	-2.94***	.66	-2.89***	.66	-2.87***	.66
Minority	4.45***	1.05	4.52***	1.05	4.52***	1.05
Age	.12	.09	.12	.09	.11	.09
Slope Regression Parameters						
Anxiety	-.54	.57	-.60	.57	-.62	.57
Mood	-3.35***	.45	-3.44***	.45	-3.50***	.45
Other	.69	.42	.59	.42	.62	.42
Male	.59	.35	.59	.35	.61***	.36
Not Single	-1.29***	.36	-1.23***	.36	-1.19***	.36
Minority	-1.09	.57	-1.03	.57	-1.05	.56
Age	.11*	.04	.11*	.05	.11*	.05
Log Number of Sessions Regression Parameters						
Anxiety	.08**	.03	.06**	.02	.06**	.02
Mood	.09***	.02	.06***	.02	.04**	.01
Other	.06**	.02	.04*	.02	.04**	.02
Male	-.04*	.02	-.03*	.01	-.03*	.01
Not Single	-.03*	.02	-.02	.01	-.01	.01
Minority	-.06*	.03	-.04	.02	-.04*	.02
Age	-.01*	.00	.00	.00	.00	.00
Variance Components						
Intercept	376.68***	9.01	377.24***	9.01	377.76***	9.04
Slope	73.36***	3.15	72.35***	3.09	71.87***	3.22
Cov	-45.25***	3.95	-45.45***	3.92	-45.34***	3.92
Residual	102.81***	2.14	102.89***	2.14	102.89***	2.14
Class Proportions and Average Number of Sessions by Class						
	$\hat{\pi}_k$	#Sessions[a]	$\hat{\pi}_k$	#Sessions[a]	$\hat{\pi}_k$	#Sessions[a]
Class 1	1.00	5.42	.30	10.50	.31	7.35
Class 2	-	-	.70	4.12	.56	3.69
Class 3	-	-	-	-	.13	14.17
Model Fit						
AIC	242,302.08		241,444.47		241,128.46	
BIC	242,489.14		241,657.33		241,367.12	

Note. * $p < .05$; ** $p < .01$; *** $p < .001$.

[a] Intercept # of sessions is the within-class average for single, White females with adjustment disorders who are of average age.

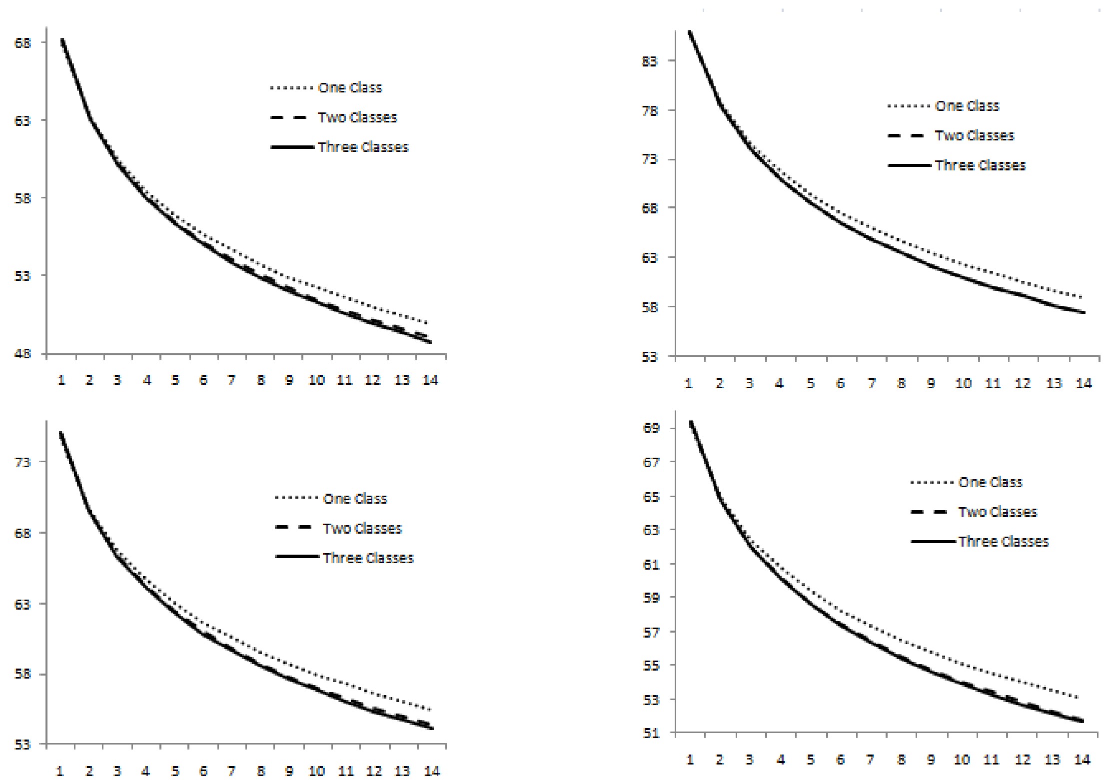


Figure 22. A Comparison of SPMM- implied trajectories for four diagnoses by number of classes. Adjustment disorder (top left); mood disorder (top right); anxiety (bottom left); other (bottom right).

In this analysis, the choice between two and three class SPMMs is not important because SPMM-implied trajectory estimates are not much different from one another; the average slope become very slightly steeper as more classes are added. Furthermore, the variance component estimates remain relatively stable from two to three classes. The main difference is between the LCM-implied trajectory (or the one class SPMM) and both of the SPMM-implied trajectories with more than one class. However, even this difference is substantively very small; effect sizes (measured as the standardized difference between the two class SPMM-implied y_{ti} at time *t* and the LCM- implied y_{ti} at time *t*) are reported in Table 14 for the four diagnostic categories.

Table 14. *Standardized Difference between LCM- and SPMM-Implied OQ-45 Scores by Time and Diagnostic Category*

Session	Adjustment	Mood	Anxiety	Other
1	.01	.01	.01	.01
2	.00	-.01	-.01	-.01
3	-.01	-.02	-.01	-.02
4	-.01	-.03	-.02	-.02
5	-.02	-.04	-.02	-.03
6	-.02	-.04	-.03	-.03
7	-.02	-.04	-.03	-.03
8	-.03	-.05	-.03	-.04
9	-.03	-.05	-.03	-.04
10	-.03	-.05	-.03	-.04
11	-.03	-.05	-.04	-.04
12	-.03	-.05	-.04	-.04
13	-.03	-.05	-.04	-.04
14	-.03	-.05	-.04	-.04

Discussion of Study 3

A comparison of LCM- (or one- class SPMM-) implied OQ-45 scores with two- or three-class SPMM- implied OQ-45 scores (either using Figure 22 or Table 14) provides evidence for the robustness of the LCM results to the random slope dependent missingness that is believed to be present in the psychotherapy data. In other words, after controlling for the data that are observed, the residual dependence between dropout occasion and rate of improvement over time is very slight. Indeed, if one were to use the specified cut-off for a clinical diagnosis using the OQ-45 (a score of 63; Lambert et al., 2004), there would be virtually no difference in the expected length of treatment, regardless of whether one used LCM-based or SPMM-based model results.

The purpose of Baldwin et al.'s (2009) manuscript was to emphasize that individual differences in rates of change render average trajectories less meaningful for calculating psychotherapy pay-offs for individuals. Using a traditional pattern mixture modeling approach, the authors showed that conditional trajectories of psychological functioning appear dramatically different from one another (after accounting for the number of sessions attended; see Figure 23, bottom). Class-specific trajectory estimates are shown for the three-class SPMM that was estimated in this manuscript on the top panel in Figure 23. The class-specific estimates show a similar pattern to Baldwin et al.'s conditional trajectories. That is, psychological trajectories depend on the number of sessions attended, suggesting the presence of a severe random coefficient dependent mechanism that, in actuality, is quite weak.

Demirtas and Schafer (2003) addressed this issue, illustrating that the mere visual appearance of random coefficient dependent missingness does not necessarily imply that it exists, saying: "Relationships between *R* [the missing data patterns] and pre-drop-out responses cannot disprove the hypothesis of ignorability, which states that there is no residual relationship between *R* and the *post-drop-out responses* given the pre-drop-out values" (italics in original). They go on to say: "Evaluating the significance of *R* -terms does not test the null hypothesis of ignorability, but the null hypothesis that drop-out is merely covariate-dependent" (pp. 2557). In other words, it is not enough to show that individual differences in growth are associated with dropout occasion because this association may be explained by observed covariates (in this case, the association between number of sessions attended and improvement in psychological distress over time was mostly explained by clinical diagnosis and other demographic variables, and by baseline OQ-45 measures). The analysis presented in the present chapter shows that, *after* controlling for baseline measured variables, individual mean trajectory estimates are largely unbiased when a traditional LCM is used. Of course, this does not detract from Baldwin et al.'s (2009) ultimate conclusion, which was that individual differences in rates of improvement over time render average improvement rates less relevant. On the other hand, a service provider has no choice but to use the information available to them before the initiation of therapy to estimate how long a patient might expect to remain in therapy, or what symptom improvement might look like.

<u>Limitations.</u> Ideally, the sensitivity analysis conducted in this chapter would lead to the conclusion that patients who leave therapy earlier in the study are similar enough to the patients who stay for more sessions, conditional on their age, gender, ethnicity, diagnosis, marital status, baseline psychological functioning, and all previously observed repeated measures, that time of dropout is not clinically significantly related to future growth. That is, it would be nice to be confident in the conclusion that the LCM-implied trajectories can validly predict individuals' expected trajectories in psychotherapy, were they to remain in the study for fourteen sessions, regardless of their background characteristics. In this study, this seems like a valid conclusion. For sake of argument, however, it is important to consider other possible explanations that could have lead us to observe minor differences between the LCM- implied and SPMM-implied trajectories.

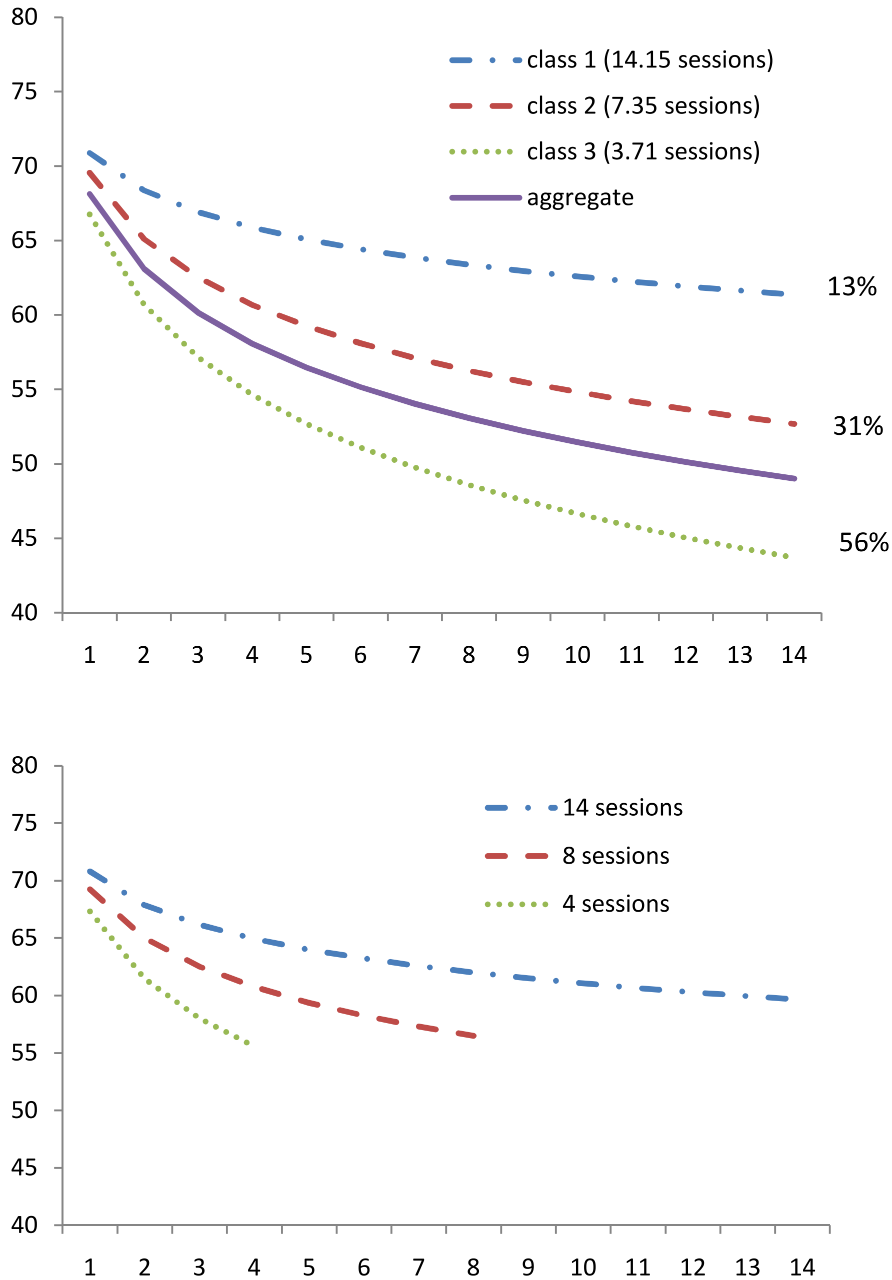


Figure 23. Model-Implied Psychotherapy Trajectories, Conditional on Number of Sessions Attended: A Comparison of the SPMM-Approach using Class-Specific Estimates (Top) and Baldwin et al.'s (2009) Traditional Pattern Mixture Approach (Bottom)[27]

---

[27] Baldwin et al.'s (2009) original analysis used a cubic growth function. The data have been re-analyzed using a log linear growth function in order to match the SPMM results.

Chapter 2 showed that there are three situations that lead to similar fixed effect estimates. The first is an MAR process, the second is a RC-MNAR-NM process whereby there is selection occurring from both sides of the random effect distribution, and the third is a OD-MNAR process. Bias in variance component estimates is also expected to be similar across all of these conditions. In other words, there is no empirical way to test whether dropout in the psychotherapy study is approximately conditionally random, whether it is due to a time-specific, outcome-dependent process (e.g., a breakthrough in therapy), or whether people who leave early are leaving due to two opposite, but non-random processes (e.g., quick improvers drop out early and people for whom psychotherapy is ineffective also drop out early). Unfortunately, all of these mechanisms are logically plausible, with the possible exception of OD-MNAR. It is also plausible that a mixture of all of these processes is operating.

If it can be assumed that a single therapy session does not provide a breakthrough that ‘cures’ patients of their psychological diagnosis, regardless of prior psychological functioning scores (i.e., if the dropout process is not OD-MNAR), then the simulation results suggest that it is safe to rely on the fixed effect estimates that were obtained in the LCM and SPMM. Reliance on variance component estimates is more uncertain, but Chapter 2 suggests that it is safe to say that the variance component estimates represent a lower bound of the true population variability. True variance components will be larger than the estimates presented here to the extent that there are non-random forces operating on both sides of the random slope distribution.

Conclusion. A primary purpose of Study 3 was to assess the feasibility of implementing the SPMM with real-world data. Implementation was virtually seamless;

the models were easily estimable using user-friendly software, and conclusions of the sensitivity analysis were fairly robust to the number of classes that were extracted. Thus, it appears that SPMM has the potential to be a useful tool for applied longitudinal researchers who are concerned about the presence of random coefficient-dependent missingness influencing their results.

As illustrated in this chapter, the SPMM should be used as a tool for the careful and thoughtful checking of the sensitivity of traditional growth model results to violations of the MAR assumption. As with all statistical tools, the SPMM should not be employed mechanically, without regard to the theoretically plausible mechanisms underlying the missing data.

CHAPTER 5

CONCLUSION

A variety of techniques for handling non-randomly missing data have been presented in the past quarter century (including major developments by Heckman, 1976; Wu & Carroll, 1986; Little, 1993; Diggle & Kenward, 1994; Roy, 2003; Lin et al., 2004, and summaries by Little, 2009, Enders, 2011, and Muthèn et al., 2011). Yet, it seems that these techniques are employed only by those who develop the methods and a handful of other applied methodologists in the social sciences (e.g., Morgan-Lopez & Fals-Stewart, 2007). Enders (2011) suggested that the slow uptake of non-ignorable missing data modeling in the social sciences has been in part due to the lack of availability of user-friendly software programs to implement these models. Muthèn et al. (2011) demonstrated how to implement a variety of missing data models that can be estimated in Mplus software. This demonstration appears in the first volume *Psychological Methods* for 2011, which contains a series of articles drawing attention to the problem of non-randomly missing data in psychological research. Thus, it appears that new-found attention to non-randomly missing data reflects the current zeitgeist of psychological research.

There may be a second reason for the reluctance on the part of applied researchers to implement models for handling non-randomly missing data: skepticism about the validity of results obtained by these models. Indeed, just as there have been numerous papers promoting methodological developments for handling missing data (particularly in the biostatistics literature), several papers have pointed out shortcomings of these models (e.g., Winship & Mare, 1992; Kenward, 1998; Demirtas & Schafer, 2003; Molenberghs, Beunckens, & Sotto, 2008), and for good reason. There is no question that every model for handling non-randomly missing data relies on untestable assumptions.

The SPMM, in particular, makes the following assumptions: 1) that OD-MNAR is not present, 2) that the missing data indicators are adequate to summarize the information necessary to account for nonignorability of the missing data process, 3) that conditional independence exists between the missing data indicators and the repeated measures (conditional on the latent classes), and 4) that it is meaningful to aggregate across missingness patterns to make inferences for the whole population.

What is less obvious, perhaps, is that the LCM (and similar commonly implemented techniques for longitudinal data analysis) also relies on an untestable assumption that missing data are MAR. In many applications, this assumption may be less tenable than those underlying SPMM or other models for MNAR data. The LCM is therefore not a justifiable modeling choice when MNAR missingness is possibly present. The problem with non-randomly missing data lies in its own nature, and not in the models used to handle it. As a number of methodologists have pointed out, the healthiest way to handle missing data is through sensitivity analyses with full awareness of the assumptions and limitations inherent in various models (e.g., Little, 1994; Verbeke, Molenberghs, Thijs, Lesaffre, & Kenward, 2001; Enders, 2011).

Beyond knowing the theoretical limitations of MNAR models, it is important to understand the practical limitations of the models under real-world data conditions. This is one of the main contributions of the present manuscript. Chapter 2 expanded Morgan-Lopez and Fals-Stewart's (2008) finding that latent mixture models work well with latent class dependent missingness to show that SPMMs also work well with a random coefficient dependent missingness that depends on latent continua, not just on latent classes. That is, this is the first research conducted that shows that the SPMM can ameliorate bias due to an MNAR process where the model provides an approximation (rather than literal embodiment) of this process. As expected, the approximation is best with random coefficient dependence missingness, but is insufficient with OD-MNAR. Additionally, the model has some difficulty recovering variance components when non-random selection operates on both ends of the random effect distribution. Encouragingly, the first study showed that there is no substantial downside to estimating a SPMM even if data are randomly missing. Finally, Chapter 2 showed that it is possible, and even desirable, to implement a more computationally feasible version of the SPMM by using a single summary indicator to represent the missing data, rather than using *t* binary missing data indicators.

Chapter 3 showed that the SPMM works better with longer longitudinal studies (i.e., studies that collect more repeated measures from participants), and the model works much better with erratically missing data than with the dropout mechanism that was tested, although this effect declines substantially as observation length increases. Generalizing from Chapter 3, it is prudent to conclude that the SPMM will perform better when repeated measures exhibit strong communality from the growth factors than when the observed indicators have low reliability. Chapter 3 showed that the SPMM is relatively robust to the proportion of missing data, controlling for MNAR severity. Thus, Chapter 3 lays out some clear guidelines for researchers considering whether the SPMM is an appropriate choice for handling suspected random coefficient dependent missingness in their data. Researchers with brief panel designs, and particularly those whose participants drop out of the study completely (rather than returning to the study erratically) might consider another choice (e.g., traditional pattern mixture models), whereas researchers with longer follow-ups, particularly those in which participants provide information throughout much of the span of the study (e.g., an experience sampling design), will probably obtain quite accurate results if they rely on the SPMM to handle missing data.

One shortcoming of simulation studies is that they assess model performance with data that are generated using relatively simple models, and with a population model that is already known prior to data analysis. Chapter 4 demonstrated the implementation of a sensitivity analysis of LCM results with a SPMM in a psychotherapy dataset in which random slope dependent missingness was suspected. The analysis suggested that the guidelines based on simulation results from Chapters 2 and 3 are generally easy to follow, and that the model is straightforward to implement with real data. Further, Chapter 4 highlighted the point that random coefficient dependent missingness is not detectable by visual inspection, so sensitivity analyses (specifically with a model like the SPMM) are necessary for evaluating whether random coefficient dependent missingness is present.

Limitations and Future Directions

As a matter of practicality, simulation studies are always limited in scope. The most pressing factors to consider were varied in Chapters 2 and 3, while other factors were held constant or limited in complexity. The most obvious limitation of the simulation studies presented here is that the generating growth model was linear in form. It is possible, and even likely, that the SPMM will experience more difficulty efficiently accounting for random coefficient dependent missingness when the number of growth factors increases. For a related model, the semi-parametric growth model (SPGM, Nagin, 1999), Sterba, Baldasaro, and Bauer (2010) found that the approximation of variance components declines as the number of latent continua increases. Unlike the SPMM, the SPGM does not allow within-class variability. Therefore, the SPMM might be more robust to the addition of growth factors than the SPGM.

A second limitation of the simulation studies presented here is that, although factor score determinacy was considered as a factor, the relative contribution of residual variance to the repeated measures was fairly low. The residual variances used in the simulations were based on the real data analyzed in Chapter 4; however, Chapter 4 utilized repeated measures that were scale scores based on 45 items, and so they were probably more reliable than most measures used in psychology. Increased residual variation in the repeated measures would probably decrease the approximating power of the SPMM, impeding its ability to quickly approach unbiased parameter estimates.

The sample size used in Chapters 2 and 3 was on the small side, and it did not align with the large sample size in Chapter 4. This misalignment was brought to bear when the AIC and BIC continued decreasing in magnitude in Chapter 4, beyond the point when aggregate parameter estimates had stopped changing. This phenomenon was not observed in Chapters 2 and 3 because the AIC and BIC tend to prefer fewer classes when sample sizes are smaller. In practice, this issue is of little importance because, as Chapter

3 showed, the AIC and BIC often overestimate the number of classes that should be used, even with a small level 2 sample size. A more careful strategy is advised; one that involves looking at standard errors and parameter change in addition to information criteria across SPMM solutions with different numbers of classes. In part, a small level 2 sample size was used in Chapters 2 and 3 in order to provide a more stringent test of the SPMM. The model should only perform better with a larger sample size.

Future research on SPMM performance should emphasize more complex models, both with respect to models of growth and with respect to missing data mechanisms. In addition, future work should compare performance of SPMM with other types of models for random coefficient dependent missingness. For instance, it would be valuable to compare performance of the SPMM with traditional PMMs when a small number of repeated measures are present, and to compare the SPMM with a parametric selection / shared parameter model in the presence of dropout. It will also be important to consider potential difficulties that may arise when categorical data are present.

The most interesting future directions will involve thoughtful, real-world applications of SPMM across a range of contexts. Hopefully, the increasing awareness of MNAR and its implications will cause researchers to stop ignoring non-ignorable missing data and to make use of the many MNAR modeling approaches that now exist. The practice of regularly conducting sensitivity analyses for missing data assumptions should be encouraged by those who engage in manuscript review, and it should be enforced by journal editors.

Appendix: Raw, Relative, and Standardized Bias by Model in Study 1

In the tables below, three measures of bias are reported. The formula for Standardized Bias is shown in Equation 2.9. Raw Bias is measured as follows:

$$\text{Raw Bias} = \frac{\sum_{j=1}^{N} \hat{\theta}_j}{N} - \theta .$$

Relative Bias scales the Raw Bias by the magnitude of the parameter:

$$\text{Relative Bias} = \frac{\text{Raw Bias}}{\theta} .$$

Bias in LCM-Generated Parameter Estimates by Missingness Mechanism

Generating Mechanism/ Parameter	Raw Bias	Relative Bias (%)	Standardized Bias (%)
MAR			
Conditional Intercept (α_0)	.09	1.30	4.89
Conditional Slope (α_1)	0.00	0.00	0.00
Residual Intercept Variance (ψ_{00})	-3.84	-1.02	-10.30

Generating Mechanism/ Parameter	Raw Bias	Relative Bias (%)	Standardized Bias (%)
Residual Slope Variance (ψ_{11})	-.11	-1.43	-9.82
Covariance (ψ_{01})	.11	-1.06	2.28
Intercept Regression (γ_0)	-.05	-.50	-1.89
Slope Regression (γ_1)	.01	-.89	2.22
SPMM-Consistent			
Conditional Intercept (α_0)	1.46	2.12	**76.84**
Conditional Slope (α_1)	-.45	**18.00**	**-128.57**
Residual Intercept Variance (ψ_{00})	-1.25	-.33	-2.87
Residual Slope Variance (ψ_{11})	-.90	-6.61	**-60.81**
Covariance (ψ_{01})	1.72	**-16.57**	27.79
Intercept Regression (γ_0)	-.05	-.50	-1.85
Slope Regression (γ_1)	.03	-2.66	-1.85
RC-MNAR-M			
Conditional Intercept (α_0)	3.1	4.49	1.63
Conditional Slope (α_1)	-1.01	**40.40**	**-404.00**
Residual Intercept Variance (ψ_{00})	-8.78	-2.34	-21.22
Residual Slope Variance (ψ_{11})	-2.72	**-37.88**	**-335.80**
Covariance (ψ_{01})	5.48	**-52.79**	**129.86**
Intercept Regression (γ_0)	.18	1.80	6.92
Slope Regression (γ_1)	-.01	.89	-2.70

Generating Mechanism/ Parameter	Raw Bias	Relative Bias (%)	Standardized Bias (%)
RC-MNAR-NM			
Conditional Intercept (α_0)		.58	22.60
Conditional Slope (α_1)		3.60	-37.50
Residual Intercept Variance (ψ_{00})	-10.89	-2.90	-29.54
Residual Slope Variance (ψ_{11})	-2.11	**-29.39**	**-270.51**
Covariance (ψ_{01})	5.18	**-49.90**	**130.15**
Intercept Regression (γ_0)	.08	.80	3.28
Slope Regression (γ_1)	.03	2.65	-8.82
OD-MNAR			
Conditional Intercept (α_0)	-.04	-3.99	**-152.78**
Conditional Slope (α_1)	.02	-2.80	28.00
Residual Intercept Variance (ψ_{00})	-33.90	-9.04	**-85.80**
Residual Slope Variance (ψ_{11})	-.48	-6.67	**-55.17**
Covariance (ψ_{01})	.69	-6.65	14.87
Intercept Regression (γ_0)	-.39	-3.90	-15.66
Slope Regression (γ_1)	.03	-2.66	8.33

Note. MAR = Missing at random; OD-MNAR = Missing not at random due to outcome dependent mechanism; RC-MNAR-M = Monotonic random coefficient dependent mechanism; RC-MNAR-NM = Nonmonotonic random coefficient dependent mechanism Values that exceed arbitrary thresholds for 'acceptable' levels of bias (Relative Bias > .10 or < -.10 and Standardized Bias > 40 or <-.49) are bolded

Bias in Binary SPMM-Generated Parameter Estimates by Missingness Mechanism

Generating Mechanism/ Parameter	Best AIC			Best BIC		
	Raw Bias	Relative Bias (%)	Standardized Bias (%)	Raw Bias	Relative Bias (%)	Standardized Bias (%)
MAR						
Conditional Intercept (α_0)	.01	.01	.55	0.00	0.00	0.00
Conditional Slope (α_1)	0.00	0.00	0.00	.01	-.40	2.86
Residual Intercept Variance (ψ_{00})	-6.15	-1.64	-16.49	-6.01	-1.60	-15.84
Residual Slope Variance (ψ_{11})	-.19	-2.65	-17.76	-.18	-2.51	-16.67
Covariance (ψ_{01})	.53	-5.11	10.73	.45	-4.34	8.98
Intercept Regression (γ_0)	0.00	0.00	0.00	.02	.20	.75
Slope Regression (γ_1)	.02	-1.77	4.26	.01	-.89	2.13
SPMM Consistent						
Conditional Intercept (α_0)	0.03	.04	1.60	.20	.29	10.53
Conditional Slope (α_1)	0.04	-1.60	11.43	-.02	.80	-5.56
Residual Intercept Variance (ψ_{00})	-2.49	-.66	-5.88	-2.83	-.76	-6.68
Residual Slope Variance (ψ_{11})	-0.04	-.29	-2.48	-.31	-2.28	-19.02
Covariance (ψ_{01})	0.16	-1.54	2.45	.71	-6.84	10.99

	Best AIC			Best BIC		
Generating Mechanism/ Parameter	Raw Bias	Relative Bias (%)	Standardized Bias (%)	Raw Bias	Relative Bias (%)	Standardized Bias (%)
Intercept Regression (γ_0)	-0.03	-.30	-1.14	-.04	-.40	-1.50
Slope Regression (γ_1)	0	0.00	0.00	0.00	0.00	0.00
RC-MNAR-M						
Conditional Intercept (α_0)	.33	.48	15.57	.24	.35	10.91
Conditional Slope (α_1)	-.14	5.60	-37.84	-.10	4.00	-25.64
Residual Intercept Variance (ψ_{00})	-3.53	-.94	-6.81	.13	.03	2.06
Residual Slope Variance (ψ_{11})	-.75	**-10.45**	**-40.54**	-.49	-6.83	18.49
Covariance (ψ_{01})	.74	-7.13	8.72	-.12	1.16	-1.00
Intercept Regression (γ_0)	.22	2.20	8.33	.20	2.00	7.49
Slope Regression (γ_1)	-.01	.89	-2.00	0.00	0.00	0.00
RC-MNAR-NM						
Conditional Intercept (α_0)	.05	.07	2.76	.05	.07	2.76
Conditional Slope (α_1)	-.01	.40	-3.57	-.01	.40	-3.57
Residual Intercept Variance (ψ_{00})	-9.08	-2.42	-24.22	-9.35	-2.49	-24.96

Generating Mechanism/ Parameter	Best AIC Raw Bias	Relative Bias (%)	Standardized Bias (%)	Best BIC Raw Bias	Relative Bias (%)	Standardized Bias (%)
Residual Slope Variance (ψ_{11})	-1.38	**-19.22**	**-118.97**	-1.68	**-23.40**	**-166.34**
Covariance (ψ_{01})	3.31	**-31.89**	**71.03**	4.00	**-38.54**	**90.70**
Intercept Regression (γ_0)	.13	1.30	5.20	.10	1.00	4.08
Slope Regression (γ_1)	-.03	2.65	-9.09	-.03	2.65	-9.09
OD MNAR						
Conditional Intercept (α_0)	-2.45	-3.55	**-126.29**	-2.59	-3.75	**-138.50**
Conditional Slope (α_1)	.06	-2.40	23.08	.08	-3.20	32.00
Residual Intercept Variance (ψ_{00})	-32.69	-8.72	**-79.56**	-33.19	-8.85	**-84.92**
Residual Slope Variance (ψ_{11})	-.46	-6.41	**-51.69**	-.47	-6.55	**-53.41**
Covariance (ψ_{01})	.92	-8.72	19.41	.71	-6.84	15.37
Intercept Regression (γ_0)	-.34	-3.40	-12.41	-.34	-3.40	-12.88
Slope Regression (γ_1)	.07	-6.20	18.42	.04	-3.54	10.81

Note. MAR = Missing at random; OD-MNAR = Missing not at random due to outcome dependent mechanism; RC-MNAR-M = Monotonic random coefficient dependent mechanism; RC-MNAR-NM = Nonmonotonic random coefficient dependent mechanism Values that exceed arbitrary thresholds for 'acceptable' levels of bias (Relative Bias > .10 or < -.10 and Standardized Bias > 40 or <-.49) are bolded

Bias in Summary SPMM-Generated Parameter Estimates by Missingness Mechanism

Generating Mechanism/ Parameter	Best AIC Raw Bias	Best AIC Relative Bias (%)	Best AIC Standardized Bias (%)	Best BIC Raw Bias	Best BIC Relative Bias (%)	Best BIC Standardized Bias (%)
MAR						
Conditional Intercept (α_0)	-0.06	-0.09	-3.26	-0.06	-0.04	-1.64
Conditional Slope (α_1)	-0.01	0.40	-2.94	-0.01	0.40	-2.94
Residual Intercept Variance (ψ_{00})	-6.93	-1.85	-18.14	-6.93	-1.90	-18.48
Residual Slope Variance (ψ_{11})	-0.28	-3.90	-25.45	-0.28	-3.90	-25.69
Covariance (ψ_{01})	0.67	-6.45	13.54	0.67	-4.62	9.90
Intercept Regression (γ_0)	0.26	2.60	9.70	0.26	3.80	14.13
Slope Regression (γ_1)	0	0.00	0.00	0	0.00	0.00
OD-MNAR						
Conditional Intercept (α_0)	-0.06	-3.12	**-118.13**	-0.06	-3.12	**-117.49**
Conditional Slope (α_1)	-0.01	-2.40	23.08	-0.01	-2.80	26.92
Residual Intercept Variance (ψ_{00})	-6.93	-6.55	**-60.43**	-6.93	-6.61	**-61.34**
Residual Slope Variance (ψ_{11})	-0.28	-6.41	**-52.27**	-0.28	-6.27	**-51.72**
Covariance (ψ_{01})	0.67	-8.57	18.46	0.67	-8.57	18.62

	Best AIC			Best BIC		
Generating Mechanism/ Parameter	Raw Bias	Relative Bias (%)	Standardized Bias (%)	Raw Bias	Relative Bias (%)	Standardized Bias (%)
Intercept Regression (γ_0)	0.26	-4.00	-15.33	0.26	-4.00	-15.56
Slope Regression (γ_1)	0	-0.88	2.78	0	-1.77	5.56
SPMM-Consistent						
Conditional Intercept (α_0)	-0.06	0.23	8.25	-0.06	0.22	7.73
Conditional Slope (α_1)	-0.01	-2.40	17.65	-0.01	-2.40	17.65
Residual Intercept Variance (ψ_{00})	-6.93	-0.51	-4.66	-6.93	-0.44	-3.97
Residual Slope Variance (ψ_{11})	-0.28	0.22	1.85	-0.28	0.22	1.85
Covariance (ψ_{01})	0.67	2.50	-4.07	0.67	2.02	-3.30
Intercept Regression (γ_0)	0.26	-2.40	-8.89	0.26	-2.40	-8.89
Slope Regression (γ_1)	0	-3.54	9.30	0	-2.65	7.14
RC-MNAR-M						
Conditional Intercept (α_0)	-0.06	0.07	2.40	-0.06	0.01	0.48
Conditional Slope (α_1)	-0.01	2.40	-18.18	-0.01	2.40	-18.18
Residual Intercept Variance (ψ_{00})	-6.93	-2.57	-20.90	-6.93	-2.60	-21.16

		Best AIC			Best BIC	
Generating Mechanism/ Parameter	Raw Bias	Relative Bias (%)	Standardized Bias (%)	Raw Bias	Relative Bias (%)	Standardized Bias (%)
Residual Slope Variance (ψ_{11})	-0.28	-8.08	-33.53	-0.28	-8.22	-34.50
Covariance (ψ_{01})	0.67	**-10.12**	14.73	0.67	-10.12	14.75
Intercept Regression (γ_0)	0.26	1.50	5.56	0.26	1.30	4.87
Slope Regression (γ_1)	0	-1.77	5.26	0	-1.77	5.26
RC-MNAR-NM						
Conditional Intercept (α_0)	-0.06	0.22	8.29	-0.06	0.23	8.84
Conditional Slope (α_1)	-0.01	0.40	-3.57	-0.01	0.80	-7.14
Residual Intercept Variance (ψ_{00})	-6.93	-2.60	-26.52	-6.93	-2.60	-26.71
Residual Slope Variance (ψ_{11})	-0.28	**-18.52**	**-108.13**	-0.28	**-19.50**	**-113.82**
Covariance (ψ_{01})	0.67	**-34.01**	**75.11**	0.67	**-34.68**	**76.43**
Intercept Regression (γ_0)	0.26	2.40	9.64	0.26	2.40	9.64
Slope Regression (γ_1)	0	-0.88	3.03	0	-0.88	3.03

Note. MAR = Missing at random; OD-MNAR = Missing not at random due to outcome dependent mechanism; RC-MNAR-M = Monotonic random coefficient dependent mechanism; RC-MNAR-NM = Nonmonotonic random coefficient dependent mechanism Values that exceed arbitrary thresholds for 'acceptable' levels of bias (Relative Bias > .10 or < -.10 and Standardized Bias > 40 or <-.49) are bolded

References

Acito, F. & Anderson, R. D. (1986). A simulation study of factor score indeterminacy. *Journal of Marketing Research, 23,* 111-118.

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control, 19*, 716-723.

Albert, P.S. & Follman, D.A. (2009). Shared-parameter models. In G. Fitzmaurice, M. Davidian, G. Verbeke, & G. Molenberghs (Eds.) *Longitudinal Data Analysis,* (pp. 433-452). Boca Raton: Chapman & Hall/CRC Press.

Arbuckle, J.L. (1996). Full information estimation in the presence of incomplete data. In G.A. Marcoulides & R.E. Schumacker (Eds.) *Advanced Structural Equation Modeling: Issues and Techniques,* (pp. 243-277). Hillsdale, NJ: Erlbaum.

Arminger, G., Stein, P., & Wittenberg, J. (1999). Mixtures of conditional mean- and covariance structure models. *Psychometrika, 64,* 475-494.

Baldwin, S. A., Berkeljon, A., Atkins, D. C., Olsen, J., & Neilsen, S. (2009). Rates of change in naturalistic psychotherapy: Contrasting dose-effect and good-enough level models of change. *Journal of Consulting and Clinical Psychology, 77,* 203-211.

Bauer, D. J. (2003). Estimating multilevel linear models as structural equation models. *Journal of Educational and Behavioral Statistics, 28,* 135-167.

Bauer, D.J. (2007). 2004 Cattel Award Address: Observations on the use of growth mixture models in psychological research. *Multivariate Behavioral Research, 42*, 757-786.

Bauer, D.J. & Curran, P.J. (2003). Distributional assumptions of growth mixture models: Implications for over-extraction of latent trajectory classes. *Psychological Methods, 8,* 338-363.

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

Bollen, K. A. (1980). Issues in the comparative measurement of political democracy. *American Sociological Review, 45*, 370-390.

Bollen, K.A., & Curran, P.J. (2006). *Latent Curve Models: A Structural Equation Approach*. Wiley Series on Probability and Mathematical Statistics.

Bozdogan, H. (1987). Model selection and Akaike's Information Criterion (AIC): The general theory and its analytical extensions. *Psychometrika, 52*, 345-370.

Brown, W. (1910). Some experimental results in the correlation of mental abilities. *British Journal of Psychology, 3*, 296-322.

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.

Curran, P. J. (2003). Have multilevel models been structural equation models all along? *Multivariate Behavioral Research, 38,* 529-569.

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

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

Diggle, P. & Kenward, M. G. (1994). Informative drop-out in longitudinal data analysis. *Journal of the Royal Statistical Society. Series C (Applied Statistics), 43,* 49-93.

Dolan, C. V. & van der Mass, H. L. J. (1998). Fitting multivariate normal finite mixtures subject to structural equation modeling. *Psychometrika, 63,* 227-253.

Eggleston, E. P., Laub, J. J. & Sampson, R. J. (2004). Methodological sensitivities to latent class analysis of long-term criminal trajectories. *Journal of Quantitative Criminology*, *20*, 1-26.

Enders, C.K. (2001) A primer on the use of maximum likelihood algorithms available for use with missing data. *SEM*, *8,* 128-141.

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

Green, B. F. (1976). On the factor score controversy. *Psychometrika, 41*, 263-266.

Grice, J. W. (2001). Computing and evaluating factor scores. *Psychological Methods, 6*, 430-450.

Guttman, L. (1955). The determinacy of factor score matrices, with implications for five other basic problems of common-factor theory. *British Journal of Statistical Psychology, 8*, 65-81.

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

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

Heckman, J. & Singer, B. (1984). A method for minimizing the impact of distributional assumptions in econometric models for duration data. *Econometrica, 52,* 271-320.

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

Hipp, J.R. & Bauer, D.J. (2006). Local solutions in the estimation of growth mixture models. *Psychological Methods, 11,* 36-53.

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

Jedidi, K., Jagpal, H.S., & DeSarbo, W.S. (1997). Finite-mixture structural equation models for response-based segmentation and unobserved heterogeneity. *Marketing Science, 16,* 39-59.

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

Lambert, M.J., Morton, .J., Hatfield, D., Harmon, C., Hamilton, S., Reid, R.C., et al. (2004). *Administration and scoring manual for the OQ-45.2.* Orem, UT: American Professional Credentialing Services.

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

Litière, S., Alonso, A., & Molenberghs, G. (2007). The impact of a misspecified random-effects distribution on the estimation and the performance of inferential procedures in generalized linear mixed models. *Statistics in Medicine, 27,* 3125-3144.

Little, R.J.A. (1993). Pattern-mixture models for multivariate incomplete data. *Journal of the American Statistical Association, 88*, 125-134.

Little, R.J.A. (1994). A class of pattern-mixture models for normal missing data. *Biometrika, 81*, 471-483.

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

Little, R.J.A. (2009). Selection and pattern-mixture models. In G. Fitzmaurice, M. Davidian, G. Verbeke, & G. Molenberghs (Eds.) *Longitudinal Data Analysis,* (pp. 409-432). Boca Raton: Chapman & Hall/CRC Press.

Lo, Y., Mendell, N. R., Rubin, D. B. (2001). Testing the number of components in a normal mixture. *Biometrika, 88*, 767-778.

Lubke, G. & Muthén, B. O. (2007). Performance of factor mixture models as a function of model size, covariate effects, and class-specific parameters. *Structural Equation Modeling, 14,* 26-47.

Lubke, G. & Neale, M. C. (2006). Distinguishing between latent classes and continuous factors: Resolution by Maximum Likelihood? *Multivariate Behavioral Research, 41*, 499-532.

Maxwell, S.E. & Yang, M. (2010). Estimation of treatment effects in randomized longitudinal designs with different types of non-ignorable dropout. Paper presented at the annual meeting of the *Society for Multivariate and Experimental Psychology* in Atlanta, GA.

McArdle, J.J., & Epstein, D. (1987). Latent growth curves within developmental structural equation models. *Child Development*, *58*(1), 110-133.

McArdle, J.J. & Hamagami, F. (1992). Modeling incomplete longitudinal data and cross-sectional data using latent structural models. *Experimental Aging Research*, *18*, 145-166.

McLachlan, G.J. & Peel, D. (2000). *Finite Mixture Models.* Wiley Series on Probability and Mathematical Statistics.

Mehta, P. D. & West, S. G. (2000). Putting the individual back into individual growth curves. *Psychological Methods*, *5*, 23-43.

Meredith, W., & Tisak, J. (1990). Latent curve analysis. *Psychometrika*, *55*(1), 107-122.

Molenberghs, G., Beunckens, C., & Sotto, C. (2008). Every missing not at random model has got a missing at random counterpart with equal fit. *Journal of the Royal Statistical Society*, *70*, 371-388.

Morgan-Lopez, A.A. & Fals-Stewart, W. (2007). Analytic methods for modeling longitudinal data from rolling therapy groups with membership turnover. *Journal of Consulting and Clinical Psychology, 75*, 580-593.

Morgan-Lopez, A.A. & Fals-Stewart (2008). Consequences of misspecifying the number of latent treatment attendance classes in modeling group membership turnover within ecologically valid behavioral treatment trials. *Journal of Substance Abuse Treatment, 35*, 396-409.

Mulaik, S. A. (1976). Comments on "the measurement of factorial indeterminacy." *Psychometrika*, *41*, 249-262.

Mulaik, S.A. & McDonald, R.P. (1978). The effect of additional variables on factor indeterminacy in models with a single common factor. *Psychometrika, 43*, 177-192.

Muthén, B. O. (2003). Statistical and substantive checking in growth mixture modeling: Comment on Bauer and Curran (2003). *Psychological Methods, 8*, 369-377.

Muthén, B.O. (2001). Latent variable mixture modeling. In G.A. Marcoulides & R.E. Schumacker (Eds.) *New Developments and Techniques in Structural Equation Modeling (*pp. 1-34). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

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

Nagin, D.S. (1999). Analyzing developmental trajectories: A semiparametric, group-based approach. *Psychological Methods, 4*, 139-157.

Raykov, T. & Marcoulides, G. A.0 (2004). Using the Delta Method for approximate interval estimation of parameter functions in SEM. *Structural Equation Modeling*, *11*, 621-637.

Rose, N., von Davier, M., & Xu, X. (2010). *Modeling Non-Ignorable Missing Data with Item Response Theory*. Paper presented at the 75th annual meeting of the International Psychometric Society in Athens, GA.

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

Roy, J. (2007). Latent class models and their application to missing-data patterns in longitudinal studies. *Statistical Methods in Medical Research, 16*, 441-456.

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

Rubin, D.B. (2004) *Multiple Imputation for Nonresponse in Surveys*. J. Wiley & Sons, New York.

Sampson, R. J., Laub, J. H. & Eggleston, E, P. (2004). On the robustness and validity of groups. *Journal of Quantitative Criminology, 1,* 37-42.

Schafer, J.L. (2003). Multiple imputation: A primer. *Statistical Methods in Medical Research, 8*, 3-15.

Schwarz, G.E. (1978). Estimating the dimension of a model. *Annals of Statistics, 6*, 461-464.

Sclove, S. L. (1987). Application of model-selection criteria to some problems in multivariate analysis. *Psychometrika, 52,* 333-343.

Singer, J. D. & Willett, J. B. (2003). *Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence*. Oxford University Press.

Snell, M.N., Mallinckrodt, B., Hill, R.D., & Lambert, M.J. (2001). Predicting counseling center clients' response to counseling: A 1-year follow-up. *Journal of Counseling Psychology, 48,* 463-473.

Spearman, C. (1910). Correlation calculated from faulty data. *British Journal of Psychology, 3*, 271-295.

Sterba, S.K. (2010). Recovery of predictor relationships via semiparametric and parametric growth models under misspecification. Ph.D. dissertation, The University of North Carolina at Chapel Hill, United States—North Carolina. Retrieved October 4, 2010, from Dissertation & Theses @ University of North Carolina at Chapel Hill. (Publication No. AAT 3408836).

Sterba, S. K., Baldasaro, R. E., & Bauer, D. J. (2010). When approximation error exceeds specification error: A comparison of semi-parametric group-based and parametric trajectory models. Unpublished manuscript. (Under review at *Psychological Methods*).

Tofighi, D. & Enders, C. K. (2007). Identifying the correct number of classes in a growth mixture model. In G. R. Hancock (Ed.), *Mixture Models in Latent Variable Research* (pp. 317 – 341). Greenwich, CG: Information Age.

Tofighi, D. & Enders, C. K. (2008). The impact of misspecifying class-specific residual variance in growth mixture models. *Structural Equation Modeling, 15*, 75-95.

Tsonaka, R., Verbeke, G., & Lesaffre, E. (2009). A semi-parametric shared-parameter model to handle nonmonotone nonignorable missingness. *Biometrics, 65*, 81-87.

Verbeke, G. & LeSaffre, E. (1996). A linear mixed-effects model with heterogeneity in the random-effects population. *Journal of the American Statistical Association, 91*, 217-221.

Verbeke, G. & Lesaffre, E. (1997). The effect of misspecifying the random-effects distribution in linear mixed models for longitudinal data. *Computational Statistics & Data Analysis, 23*, 541-556.

Verbeke, G., Molenberghs, G., Thijs, H., Lessafre, E., & Kenward, M. G. (2001). Sensitivity analysis for nonrandom dropout: A local influence approach. *Biometrics, 57*, 7-14.

Vermunt, J.K. & van Dijk, L.A. (2001). A non-parametric random coefficient approach: The latent class regression model. *Multilevel Modeling Newsletter, 13*, 6-13.

Vonesh, E.F., Greene, T., & Schlucher, M.D. (2006). Shared parameter models for the joint analysis of longitudinal data with event times. *Statistics in Medicine*, *25*, 143-163.

Winship, C. & Mare, R.D. (1992). Models for sample selection bias. *Annual Review of Sociology, 18*, 327-350.

Wothke, W. (2000). Longitudinal and multigroup modeling with missing data. In T. D. Little, K. U. Schnabel, & L. Baumert (Eds.), *Modeling longitudinal and multilevel data* (pp. 219–240). Mahwah, NJ: Erlbaum

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.

Yang, M. & Maxwell, S.E. (2009). Abstract: Treatment effects in randomized longitudinal experiments with different types of non-ignorable dropout. *Multivariate Behavioral Research, 44,* 856.

Yung, Y. (1997). Finite mixtures in confirmatory factor-analysis models. *Psychometrika*, *62*, 297-330.