8-2024

Performance of Shared Parameter Missing Data Models for Intensive Longitudinal Data

Justin Harding
University of Northern Colorado

UNIVERSITY OF NORTHERN COLORADO

Greeley, Colorado

The Graduate School

PERFORMANCE OF SHARED PARAMETER MISSING DATA MODELS FOR INTENSIVE LONGITUDINAL DATA

A Dissertation Submitted in Partial Fulfillment
of the Requirements for the Degree of
Doctor of Philosophy

Justin Harding

College of Education and Behavioral Sciences
Department of Applied Statistics and Research Methods

August 2024

This Dissertation by: Justin Lee Harding

Entitled: *Performance of Shared Parameter Missing Data Models for Intensive Longitudinal Data*

has been approved as meeting the requirement for the Degree of Doctor of Philosophy in the College of Education and Behavioral Sciences in the Department of Applied Statistics and Research Methods

Accepted by the Doctoral Committee

____
Dr. Chai-Lin Tsai PhD., Co- Research Advisor

____
Han Yu, PhD., Co- Research Advisor

____
Dr. William Merchant PhD., Committee Member

____
Sue Hyeon Paek, PhD., Faculty Representative

Date of Dissertation Defense ____

Accepted by the Graduate School

____
Jeri-Anne Lyons, Ph.D.
Dean of the Graduate School
Associate Vice President for Research

ABSTRACT

Harding, Justin. *Performance of Shared Parameter Missing Data Models for Intensive Longitudinal Data*. Published Doctor of Philosophy dissertation, University of Northern Colorado, 2024.

Ecological Momentary Assessment (EMA) studies, also known as Intensive Longitudinal Data (ILD), involve participants that are intensively measured over time. Intermittent missing data tends to occur due to participants not responding when prompted. The high volume of assessments and intermittent nature of missingness have made some traditional longitudinal missing data methods unsuited to handle the missingness of EMA data. Two recent missing models intended for ecological momentary assessment missing data situations have emerged that jointly model the outcome and missingness, providing information about the latent trait of responding to prompts. Both models implement a shared parameter as a random effect but do so in different ways. X. Lin et al. (2018) model the missing process by using a random intercept logistic regression model for the binary missing prompt indicators. Cursio et al. (2019) model the missing process using item response theory to model responsiveness to the prompting device as a latent trait. The purpose of this study was to compare these two joint models used to handle missing data in ecological momentary assessment (EMA) studies and to evaluate their performance under different assessment and missing data scenarios. A simulation was designed to compare these two joint models under a few different assessments and percentage of missing prompts scenarios to evaluate their performance in terms of parameter estimate bias, empirical standard errors, and computation run time. Results in this missing data simulation displayed that the joint shared parameter missing data models consistently outperformed statistical software's default missing data method list-wise deletion displaying the value of these models in ILD missing data situations. The Latent Trait Shared Parameter Mixed Model (LTSPMM) performed superior in this simulation and is recommended as a missing data model in ILD studies. The results of this study provides researchers with guidance on the performance of both shared parameter missing data models under missing data conditions that might be observed in real ILD data situations.

TABLE OF CONTENTS

CHAPTER

LIST OF TABLES

TABLE

LIST OF FIGURES

FIGURE

CHAPTER I

INTRODUCTION TO THE STUDY

Missing data is a frequent problem in longitudinal studies as participants miss observations or drop out completely. The loss of information from missing data causes severe bias and reduces precision in estimation that compromises inferences and potentially bias the results. Knowing how to manage missing data will prove extremely valuable to reduce bias and lessen misleading inferences. The result of poorly informed missing data handling is often loss of statistical power, as well as biased and inefficient parameter estimates that may lead to incorrect conclusions about the nature of variable relationships in the population (Black et al., 2011). This chapter will provide an introduction on longitudinal data, missing data mechanisms (Rubin, 1976), and missing data methods for longitudinal studies. The following section will introduce ecological momentary assessment studies, their missing data situation, emerging missing data models, and concludes with research questions and study limitations.

Longitudinal studies involve repeated measurements of the same properties from the same individuals with the intentions of capturing trends over time. These studies offer the ability to study within- and between-subject variations. The fluctuations of the participants can be assessed through time-independent and time-dependent covariates. The goal of researchers is that all participants have the same number of measurements over the course of the study. However, tracking individuals over time has proven to be complicated with missing measurements or dropout occurring at any time throughout the study for a multitude of reasons. A common type of missingness in longitudinal studies is dropout. The problem with the missing responses is loss of power and efficiency resulting from larger standard errors and bias. For example, sometimes the participants have missing data for a reason and the collection of participants that stick around may not represent the populations intended. Next, I will give a brief background on longitudinal missing data models.

Flawed methods for handling missing data involve removing observations of incomplete data or by filling in a single missing value (i.e., list-wise deletion and single imputation). Deleting the missing data is a strategy that is firmly entrenched in statistical software packages and is exceedingly common in many research disciplines (Peugh & Enders, 2004). As a default option in some statistical software packages, list-wise deletion is still used today even though it requires a strict assumption about the missing data and is prone to substantial bias. In a meta-analysis of longitudinal studies from three major journals between the years 2000 to 2006, Jelicic et al. (2009) found that 82 out of 100 articles were using traditional missing data methods that are statistically problematic. The implication of not addressing missing data can lead to biased misleading inferences so properly handling missing data will help mitigate the bias while gaining a more accurate understanding of individuals over time.

A breakthrough for missing data happened when Rubin (1976) outlined a theoretical framework for missing data problems that remains in widespread use today. The missing data mechanisms were designated Missing Completely At Random (MCAR), Missing At Random (MAR) and Missing Not At Random (MNAR) which will be described in detail in chapter 2. Properly applying Rubin’s mechanisms helps the researcher to understand their missing data situation and modern longitudinal missing data methods help improve their analysis with the possibility of estimating unbiased estimators.

Longitudinal missing data methods have received considerable attention in the methodological literature during the past 50 years. Researchers applied Rubin’s mechanisms to create different longitudinal missing data models depending on the mechanism and probability of missing. For example, full information maximum likelihood (FIML) and multiple imputation (MI) are missing data methods that are useful in MAR missingness scenarios and aim to preserve key relationships among variables and better estimate the variability in the data. Selection of missing data techniques should be done with the primary goal of preserving the distributional characteristics of the variables of interest, as well as their interrelationships with other variables for the purpose of deriving valid and meaningful inferences from the available data (Schafer, 1997; Schafer & Graham, 2002). As the literature has grown and researchers have gained more knowledge about missing data, along with increased statistical software capabilities, there has been a rise in implementing missing methods in longitudinal studies. Hayati Rezvan et al. (2015) identified 103 articles in medical research that used multiple imputation in an increasing trend from the years 2008 to 2013.

There is not one perfect method to manage missing data and the literature is constantly growing with missing data techniques to help researchers for the many diverse data situations. Enders (2010) warns that missing data handling techniques are only as good as the veracity of the assumptions they rely on, so thoughtfully applying these models is always important.

Ecological Momentary Assessment

Ecological Momentary Assessments (Stone & Shiffman, 1994) are a modern version of longitudinal projects where researchers gather a high frequency of observations on participants. These types of studies are sometimes referred to as Intensive Longitudinal Data (Walls & Schafer, 2006) or experience sampling (deVries, 1992; Hektner et al., 2007; Larson &

Csikszentmihalyi, 1983). The primary goal is to study psychological and behavior events by repeatedly collecting the momentary states of the participants in their natural environment over a predetermined timeframe. The assessments capture events, behaviors, and moods in the moment or after a slight time lapse with the hope of avoiding participant recall bias. For example, the self-report assessments are setup to ask about the current or recent emotional states of the participants, rather than asking them to recall or summarize their states over longer periods of time. What ecological momentary assessments (EMA) studies have in common is the collection of assessments of subjects' current or recent states, sampled repeatedly over time, in their natural environments (Shiffman et al., 2008). The observations get real-time, real-world behaviors in the participants' daily settings warranting ecological validity. For example, if one is interested in how subjects feel in a relationship, asking them in their daily life as it unfolds will provide better information than asking them in a research clinic. EMA research ensures ecological validity by collecting data in the real world, ensuring that the data represent the full range of real-life experience (Stone et al., 2007). EMA addresses the unchanging nature of cross-sectional surveys by analyzing dynamic association over time.

The repeated sampling in EMA studies allows researchers to capture experiences and changes within-subjects over time and across contexts during their everyday life. The within-subject relationships reveal subtle and immediate effects of momentary states allowing researchers to gain a better understanding about the process of behaviors and events. Analyses of within-subject relations yield insights into the dynamic association between variables and their dependence on situational circumstances (Bolger et al., 2003). For example, to examine if people have a positive state after being physically active, measuring the within-subject states over time will help to provide context of activity. The repeated real time assessments allow time-dependent processes to expose contextual information of both within and between participant momentary states.

EMA studies often examine the sequences of experiences leading to behaviors or events. In this type of analyses, the order of events or experiences are the key focus and are explicitly studied. The degree to which an individual's emotional state at a given time point is predictive of his/her emotional state at subsequent time points (Jahng et al., 2008). For example, does high craving of a substance at one time point predict substance use at following time points? The temporal sequence enables researchers to describe and analyze events and state behaviors over periods of hours or days. This is another way that EMA can display dynamic relationships over time.

EMA procedures entail repeatedly prompting participants to complete short surveys over the course of hours, days, or weeks. In modern computerized EMA designs, participants are first ‘‘beeped’’ by a device or a beeper, Personal Digital Assistant (PDA), phone call, or text message sent to a smartphone, and then complete a brief questionnaire about what they are doing, thinking, and feeling at the moment (Hektner et al., 2007). Researchers determine the number of prompts allowing enough time to capture how behaviors and events unfold in the context of their research questions. Although it is subjective and content-dependent, these studies tend to generate large numbers of observations per participant. Shiffman et al. (2008) indicated that assessing subjects 3 to 5 times per day is common. Particular threats to repeated measures studies, and specifically those involving momentary assessment, are fatigue, forgetfulness, noncompliance, and dropout (Black et al., 2011). The random nature of the prompting mechanism combined with participants going about their daily lives inevitably leads to some participants having nonresponse on the prompts. Participants often ignore the “beep” as they are in their natural environment, not the constrained context of a research lab (Silvia et al., 2013). Unlike longitudinal studies, it is rare for participants to drop out; instead, they miss a prompt then return to the study, which is termed intermittent missingness. Typically, EMA procedures involve instant short surveys from individuals over the course of hours, days, and weeks, where relatively large numbers of measurements per subject are produced and intermittent missingness due to non-responses can be an issue (Sokolovsky et al., 2014). This leads to each participant having varying amounts of recorded measurements and non-response for a high frequency of observations. Recent meta-analyses report that the compliance rates to the prompting device across multiple disciplines ranges between seventy and eighty percent (Jones et al., 2019; Liao et al., 2016; Wen et al., 2017).

In many situations, modern longitudinal missing data methods may not be appropriate for EMA missing data. For example, a small number of unique patterns are needed for pattern mixture models to be successful. Intermittent missingness among the participants can create a considerable number of missing patterns making it more difficult to implement a pattern mixture model. The assumptions about the nature of the missing data are typically unknown in EMA studies and, in many cases, the missing data is complex and highly irregular (Cursio et al., 2019).

In recent years, two models have emerged that offer ways of analyzing the missing EMA data with intermittent response patterns. X. Lin et al. (2018) presented a shared parameter modeling approach that links the primary longitudinal outcome with informative missingness by a common set of random effects that summarize subjects’ specific traits in terms of their mean (location) and variability (scale). Cursio et al. (2019) presented a model that utilized item response theory to model responsiveness to the prompting device as a latent trait. In this situation, the latent trait is modeled jointly with a mixed model for bivariate longitudinal outcomes. To date, these two EMA missing models have been studied separately and have not been evaluated under simulated datasets in one study. The purpose of this study was to compare these two joint models used to handle missing data in ecological momentary assessment (EMA) studies and to evaluate their performance under different assessment and missing data scenarios. In this dissertation, I intend to compare the performance of these two models by simulation and the illustration of an application to real intermittently missing data by the EMA models under various prompting and missingness conditions.

Research Questions

The following list of research questions will be answered by comparing the performance of the two shared parameter missing data models.

Q1 Which model, ILD missing data models LTSPMM (Cursio et al., 2019), SPLR (X. Lin et al., 2018), or the full mixed-effect location random effects model using list-wise deletion, perform better under different combinations of number of prompts (25, 40) and intermittent missingness scenarios (20%, 30%) in terms of raw bias percentage?

Q2 Which model, ILD missing data model LTSPMM (Cursio et al., 2019), SPLR (X. Lin et al., 2018), or the full mixed-effect location random effects model using list-wise deletion perform better under different combinations of number of prompts (25, 40) and missingness scenarios (20%, 30%) in terms of empirical standard errors?

Q3 Which ILD missing data model LTSPMM (Cursio et al., 2019) or the SPLR model (X. Lin et al., 2018), performs more computationally efficient in terms of computational run time?

Limitations of Study

This study has some possible limitations. Most computational simulation models are simplified to allow for an easier understanding of a complex phenomenon. Real world data situations may incorporate many variable types with diverse distributions making the missing data more complex. Building models that are too complicated are not feasible as they can lead to difficulty demonstrating the effect that the missing data models have on the results. The specification of deciding which factors to include in the model is another limitation. The simulation model uses variables based on an empirical study focusing on certain factors that are relevant to a theory or hypothesis being studied. However, designing EMA studies is quite complex with several factors influencing each other. Thus, specifying the factors that are incorporated in this model ensure that the model is not so simplified that it would generate results of little significance.

In terms of organization for this dissertation, chapter II will provide a literature review on modern longitudinal and EMA missing data models. First, a background on missing data for longitudinal studies, which includes the longitudinal mixed-effects model, missing data patterns, missing data mechanisms, traditional and modern missing data techniques. The second part of the literature review will provide information on EMA missing data models using multiple imputation, shared parameter with item response theory and shared parameter location scale model. Chapter III will include information on the proposed method for analyzing the two EMA missing data models by simulation as well as a real data application. The results will be discussed in Chapter IV and the discussion and conclusion in Chapter V.

CHAPTER II

REVIEW OF LITERATURE

The goal of this section is to provide a theoretical overview of missing data mechanisms and common missing data techniques for longitudinal and Intensive Longitudinal Data (ILD) studies. The chapter will begin with a discussion on longitudinal missing data providing background on the linear mixed-effects model, missing data mechanisms, traditional and modern missing data models. In the following section, there will be a review of ILD, patterns of missing data, and an introduction to the linear mixed-effects location scale model. The chapter will conclude with review on three distinct types of ILD missing data models: multiple imputation missing data technique, shared parameter missing model utilizing item response theory and shared parameter missing model applying a logistic regression model.

Longitudinal and ILD research projects involve collecting repeated observations on the same individuals over time. One of the main obstacles for them is what to do when individuals miss observations or drop out completely. The loss of information from missing data can cause imbalanced data, loss of precision and in some circumstances can introduce bias that led to misinformed inferences. Understanding the association that the missing data has with the variables may lead to unbiased conclusions. Rubin (1976) and colleagues (Little & Rubin, 2002) produced a classification system to describe relationships between the probability of missing data and variables. Properly understanding the type of missing data is a fundamental part of the data analysis process that leads to better results.

Linear Mixed-Effects Model

The modeling procedure that researchers in longitudinal and ILD studies typically use to account for these data situations is the linear mixed-effects model (LMM) introduced by Laird and Ware (1982). According to Molenberghs and Verbeke (2001), the LMM has become the primary method for analysis of longitudinal data. The multilevel model includes random subject effects that account for individual fluctuations allowing researchers to assess the inter-individuals along with between person variations by relaxing the homogeneity of variance assumption from the classical models. LMM offers more flexibility in terms of repeated measures as participants can have different numbers of observations and time can be continuous rather than a fixed set of points. The covariance structure among repeated measures offers flexible specification. LMM is more flexible in term of repeated measures and does not require restrictive assumptions concerning missing data across time and the variance–covariance structure of the repeated measures (Nakai & Ke, 2011). The LMM equation is

$$Y_{it} = \boldsymbol{X_{it}^{T}\beta} + \boldsymbol{Z_{it}u_{i}} + \varepsilon_{it}. \tag{2.1}$$

In this equation, i refers to the participant and $i = 1, \ldots, N$ where N is equal to the total number of participants. The number of time points t is measured for each participant and can vary by participant $t = 1, \ldots, n_i$. The vector Y_{it} refers to collected longitudinal responses for a given participant. The design matrix for the model covariates $\boldsymbol{X_{it}}$ includes all fixed and time-dependent covariates including a column of ones for the intercept term and has the dimension $n_i \times p$. The vector $\boldsymbol{\beta}$ includes all fixed coefficients, which includes time in the model and has dimension $p \times 1$. The term Z_{it} represents the $n_i \times r$ design matrix for the random effects terms u_i, which are the random intercept u_{0i} and random slope u_{1i} (r=2) in this study. This design matrix includes a column of ones for the random intercept term u_{0i} and additional columns for covariates that are allowed to vary among participants. In most situations, the columns in Z_{it} are a subset of columns included in $\boldsymbol{X_{it}}$. The assumptions for the LMM involve validity of the model, independence of the data points, linearity of the relationship between the predictors and the response, and absence of measurement error in the predictor. The random effects u_i are $(u_0, u_1) \sim iid\ N(0, \sigma^2_\varepsilon)$ and a variance/covariance structure of

$$\underline{\underline{G}} = (\underline{u}) = \begin{pmatrix} \sigma^2_{u_0} & \sigma_{u_{01}} \\ \sigma_{u_{01}} & \sigma^2_{u_1} \end{pmatrix}, \tag{2.2}$$

ε_{it} is an independent error term with the distribution of $\varepsilon_{it} \sim iid\ N(0, \sigma^2_\varepsilon)$, and lastly ε_{it} and u_i are independent.

Missing Data Mechanisms

What follows is a brief description of the characteristics of longitudinal data situations. The outcome vector is $\boldsymbol{Y_{it}}$ and $\boldsymbol{X_{it}}$ is the fixed design matrix with i representing the subject and t represents the time point of the assessments. The participants receive multiple assessments of data collection at non-fixed times resulting in intra-subject correlation amongst the observations of each subject. Missing data can occur to the dependent and independent variables at any time for a multitude of reasons. Participants may miss full assessments. Common types of missing data for longitudinal studies are attrition or drop out. For example, the researcher could lose track of participants thus losing them in the study. As participants miss assessments, the data structure is typically unbalanced with each participant having a different number of total responses. Longitudinal studies typically involve an interest in both the within-subject and between-subject time trends. The data collected can measure both time-independent and time-dependent predictors. The results may be biased, or incorrect conclusions could be made if the proper modeling procedures are not used to manage the many diverse types of missing data situations.

Rubin (1976) had a theory that the missingness is its own variable that has a probability distribution, and thus proposed missing data mechanisms. Missing data mechanisms are descriptions of relationships between measured or unmeasured variables and the probability of missing data. Rubin perceived the data as a complete data set that includes the observed and missing values for each variable. This creates a probability distribution for each variable and a probability for the missingness, determining whether the probability of missing for the variables are related in the data set. The type of the relationship between missing values and the data provides information distinguishing the type of missing data mechanisms. Knowing about the missing data mechanisms are important because the missing information may influence the estimation of the variables. Information from the missing data mechanisms provides information about how to analyze the missing data in the data set. Rubin (1976) and his colleagues (Little & Rubin, 2002) branded the three missing data mechanisms on how to differentiate the missingness: missing completely at random, missing at random, and missing not at random.

Missing Completely At Random (MCAR) is the mechanism when missing data are completely unsystematic. The mechanism occurs when the probability of missing data on a variable X is not related to the other variables and the possible values of those variables

$$P(m_{it}|Y_{it}^{o}, Y_{it}^{m}, X_{it}) = P(m_{it}). \tag{2.3}$$

The likelihood of missing is independent of all the observed and unobserved values. The indicator is m_{it} of missingness for individual i at time point t, Y_{it}^{o} are the given observed responses, Y_{it}^{m} are the unobserved responses and X_{it} are the independent variables. When the data is MCAR, the observed data is considered a random subsample if the data had been complete. The results of the data analysis on the data with missing values have no bias but lower power than having all the information in a complete dataset. For example, students could unexpectedly be absent on the day that achievement exams are taken and missing the exam has nothing to do with the scores they would have achieved. MCAR is a restrictive assumption that assumes missingness is entirely unrelated to the data set, making it a rare occurrence in longitudinal datasets. Jelicic et al. (2009) state that in real-world social science applications, data that are MCAR are the least likely.

Data are Missing At Random (MAR) when the probability of missing data on an outcome variable Y_{it} is related to some other measured variable in the model but not to the values of Y_{it}. There exists a systematic relationship between one or more measured variable and the probability of missing data. Suppose m_{it} is the indicator of missingness for individual i at time point t, Y_{it}^{o} are the given observed responses, Y_{it}^{m} are the unobserved responses and X_{it} are the independent variables. The model

$$P(m_{it}|Y_{it}^{o}, Y_{it}^{m}, X_{it}) = P(m_{it}|Y_{it}^{o}, X_{it}) \qquad (2.4)$$

displays the likelihood of missing is independent of unobserved values but can depend on observed outcome and independent variables. The missingness is believed to be a by-product of other variables that are measured in the study. For example, if some of the students were judged exempt for the achievement exam because of their strong class performance. The likelihood of missing data clearly depends on $X =$ performance, which gives bias to the population. MAR is a much more flexible missing data mechanism than MCAR.

Data are Missing Not At Random (MNAR) when the probability of missing data on a variable can be a function of unobserved values (Rubin, 1976). Suppose m_{it} is the indicator of missingness for individual i at time point t, Y_{it}^{o} are the given observed responses, Y_{it}^{m} are the unobserved responses and X_{it} are the independent variables. The probability of the missing data is determined by the variable that is missing

$$P(m_{it}|Y_{it}^{o}, Y_{it}^{m}, X_{it}) = P(m_{it}|Y_{it}^{o}, Y_{it}^{m}, X_{it}). \qquad (2.5)$$

In this situation, the likelihood of missing can depend on unobserved values. When missingness are thought to be MNAR, it cannot be ignored and is the most problematic of all the missing data mechanisms. In this situation, the problem of missing data should not be ignored as individuals in the study have chosen not to respond or dropped out completely which are related to the variables and values in the study. For example, if achievement exams are scheduled in a way that creates conflicts with struggling students' schedules that makes them unable to attend. The likelihood of missing data depends on something we did not observe and can never be determined by the data resulting in a biased sample. In these situations, researchers cannot make proper conclusions because of the unobserved responses and need to find a missing data method remedy. While the Little MCAR test (Little, 1988) can eliminate if the missing data mechanism is MCAR, the practical problem for researchers is confirming the missingness between MAR and MNAR, which cannot be tested. When missingness is non-ignorable, it means that we cannot predict future responses, conditional on past-observed responses; instead, we need to incorporate a model for the missingness mechanism (Nakai & Ke, 2011).

Longitudinal Missing Data Methods

Traditional missing data methods need to be examined to highlight the strengths of modern missing data methods. The two types of traditional methods in the literature are reduction and single augmentation methods. Reduction methods encompass removing cases with incomplete data. Augmentation methods are implemented by filling in values for the missing incomplete data with a single value. Both methods are problematic even if the missingness is MCAR and not recommended if the probability of missing data mechanisms is MAR or MNAR.

Reduction Methods

Reduction methods are amongst the most commonly used missing data methods in literature (Lang & Little, 2018; Peugh & Enders, 2004). The two techniques are list-wise deletion and pairwise deletion. List-wise deletion completely removes any case that has one or more missing values from the analysis. Pairwise deletion maximizes all data available by an analysis-by-analysis basis. A correlation is calculated using all cases for which data is available computing each element in a correlation matrix. The method attempts to use as much of the data as possible and tends to have higher power than list-wise deletion. An advantage of reduction methods is that they are standard options in statistical software packages and are easy to implement (Peugh & Enders, 2004; Enders, 2010). However, discarding data reduces power and wastes information about variables in the data set. Reduction methods have been researched extensively producing biased parameters when the MCAR assumption does not hold and are found to be some of the worst possible missing data methods to apply (Bodner, 2006; Enders & Bandalos, 2001; Wilkinson, 1999).

Single Augmentation Methods

Augmentations methods involve generating a value that fills in data to the missing values prior to analysis. Filling in all the missing values produces a complete dataset providing convenience of analysis and making use of all the collected data. The problem arises when generating single imputation values. The disadvantage is that imputing a single value treats the value as real data and cannot reflect sampling variability under one model. Single imputations techniques underestimate the standard errors. Augmentation methods that replace missing values with a single value is a bad missing data process.

Last observation carried forward (LOCF) replaces missing values by the last observed value from the same participant-preceding dropout. The final available observation will fill in all the subsequent missing values after the participant leaves the study. For example, if a participant drops out after the fourth week of a six-week study, the week four score fills in the remaining waves of data. This method artificially populates the sample size by an implicit assumption that participants would have maintained their last observed levels on all variables. LOCF reduces variability underestimating the standard errors in the outcome and can seriously compromise a study's inferences and lead to highly invalid conclusions (Enders, 2010; van Buuren, 2011). The bias in LOCF studies is difficult to predict and are likely to produce misleading parameter estimates even when the probability of missing is MCAR (Molenberghs et al., 2004).

Mean imputation fills in all missing values with the average value of the available observed cases. Like LOCF, mean imputation is convenient for analysis because it produces a complete data set inflating power. The downside of this method is that the mean is biased, and variance is reduced (Donders et al., 2006). This approach severely biases the resulting parameter estimates, even when the data are MCAR. Nakai and Ke (2011) emphasize that mean imputation is an unaccepted method.

Conditional mean imputation (also known as regression imputation) builds a regression model for all observed values and fills in the missing values from the fitted estimates of the model. The key idea to this approach uses information from the complete variables to fill in all the incomplete variables into one complete data set. Conditional mean imputation will impute the data with a perfectly correlated score. The problem is that it does not account for variability in the unobserved value and produce biased means (Greenland & Finkle, 1995; Olinsky et al., 2003; van Ginkel et al., 2020). The result will overestimate correlations and R^2 statistics for all missing data scenarios.

Missing at Random Missing Data Models

The two types of missing data techniques researchers turn to when the probability of missing data is considered MAR are Full Information Maximum Likelihood (FIML) and Multiple Imputation (MI). FIML uses all available information while MI fills in all the missing values. The two methods are asymptotically equivalent and tend to produce similar results.

Full Information Maximum Likelihood

FIML (Anderson, 1957) is a maximum likelihood estimator that is robust to ignorable item nonresponse. This method is easy to implement and available in statistical software packages, making it an attractive tool to manage missing data on longitudinal datasets. FIML is a missing data technique that when used under a MAR data situation produces unbiased parameter estimates (Enders, 2010). The theory behind FIML employs the probability density function of a multivariate distribution. The estimation procedure involves continually looking for the population parameters that represent the best fit for the data. FIML tend to be more powerful than traditional data techniques because no data are thrown out (Baraldi & Enders, 2010). FIML uses different individual log-likelihoods that can vary with each participant utilizing only the variables and parameters that have observed data from that participant. It oversees different missing data patterns for each participant. The main goal of maximum likelihood estimation is to maximize the parameter estimates based on the log-likelihood of the data available. Under the assumption of multivariate normality, the log likelihood function of each participant i is:

$$logL_i = K_i - \frac{1}{2}\log|\boldsymbol{\Sigma}_i| - \frac{1}{2}(\boldsymbol{y}_i - \boldsymbol{\mu}_i)'\,\boldsymbol{\Sigma}_i^{-1}\left(\underline{\boldsymbol{y}}_i - \boldsymbol{\mu}_i\right), \tag{2.6}$$

where x_i is the vector of observed values for case i, K_i is a constant that is determined by the number of observed variables for case i, and μ and Σ are, respectively, the mean vector and covariance matrix that are to be estimated (Enders, 2001). By summing the N case-wise likelihood functions for the overall discrepancy value are as follows:

$$logL(\beta, \sigma^2) = \sum_{i=0}^{N} logL_i. \tag{2.7}$$

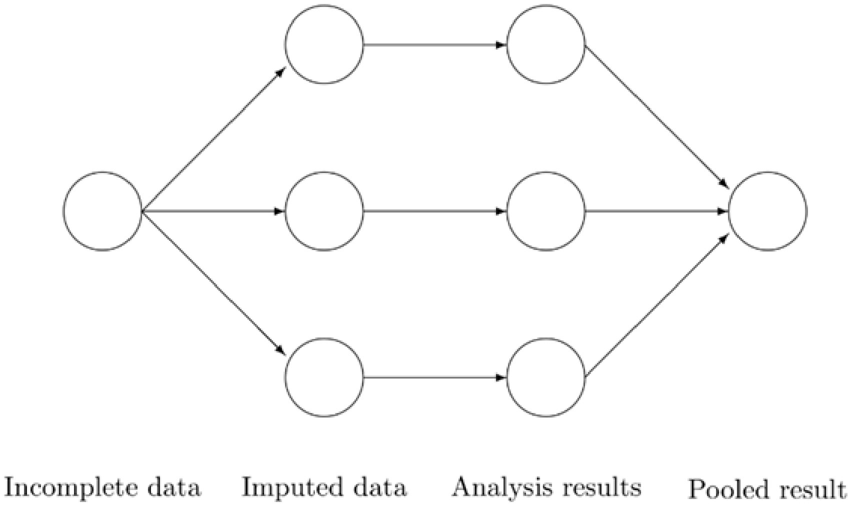
FIML is a tool for missing data that is MCAR or MAR by maximizing the statistical power of the data by using the observed information to calculate parameter estimates. On the other hand, there are consequences if the researcher does not correctly define the missing data mechanism leading to potentially seriously biased estimates if the data is MNAR. FIML, similar to most modern missing data methods, was not applied often to missing data in the early 2000s (Peng et al., 2006). The use of FIML for a missing data model has become more relevant in longitudinal research studies as there is more familiarity with the method and improved technology. For example, Dong and Peng (2013) reviewed quantitative studies published in the Journal of Educational Psychology, from 2009 to 2010 and found 12 of 46 (26%) implemented FIML as a missing data technique.

Multiple Imputation

Multiple Imputation (Rubin, 1987) is an alternative method to FIML that has the same assumptions that the missing data is MAR. Multiple imputation enables all participants to be included in the analysis and may reduce bias and improve precision of the parameter estimates compared to a complete case analysis (De Silva et al., 2017). The imputation phase generates researcher-defined copies of the data set each of which contain different estimates of the missing values. Multiple imputation analysis consists of three steps: the imputation phase, the analysis phase, and the pooling phase. In Figure 1, van Buuren, (2018) displays a scheme of the main steps in multiple imputation.

Figure 1

Diagram for the Multiple Imputation Process



The procedure for the imputation phase has two steps to make each imputation copy: I-step and P-step. The goal in this phase is to generate a number of researchers defined complete data sets with different estimates for the missing data. The I-step builds a set of regression equations from estimates of the mean vector and the covariance matrix where a comparison of the missing data pattern to the completed data predicts incomplete variables for that pattern. Using the observed data in these equations creates predicted scores into the missing variables. This step amounts to regression imputation as the predicted scores relate directly to the regression predicted line. Inserting a normal distributed residual term to each predicted value adds variability between imputations procedures. The following equation summarizes the I-step

$$Y_{it}^{*} \sim p(Y_{it}^{m} | Y_{it}^{o}, \theta_{it-1}^{*}) \tag{2.8}$$

where Y_{it}^{*} represents the imputed values at I-step t, Y_{it}^{m} is the missing portion of the data, Y_{it}^{o} is the observed portion of data, and θ_{it-1}^{*} denotes the mean vector and the covariance matrix from the preceding P-step.

The P-step uses Bayesian analysis to define the mean vector and covariance matrix of the posterior distribution. This step takes the estimates of the mean vector and covariance matrix from the I-step and creates new parameter values by adding a random residual term in the complete-data mean vector and covariance matrix. The goal is to sample new estimates of the mean vector and covariance matrix from the posterior distributions. The next I-step uses these new parameter values to create another set of regression coefficients. The multiple draws are made to move away from the initial values. Repeating this process until a specified number of copies are generated, give each copy unique estimates for the missing values. In this step, the following displays the distributions for

$$Y_{it}^{*} \sim N(0, 1), \tag{2.9}$$

$$u \sim N(0, \sigma_u^2), \tag{2.10}$$

$$\sigma^2 \sim \Gamma(\alpha, \beta). \tag{2.11}$$

Here is a summary of the P-step equation:

$$\theta_{it}^{*} \sim p(\theta | Y_{obs}, Y_{it}^{*}) \tag{2.12}$$

where θ_{it}^{*} denotes the simulated parameter values from P-step t, Y_{it}^{o} is the observed data, and Y_{it}^{*} contains the imputed values from the preceding I-step.

After the imputation phase creates the filled in imputed data sets, the researcher then must analyze the datasets in the analysis phase. Graham et al. (2007) simulation studies suggest that 20 imputations are sufficient for many realistic situations and increasing the number of imputations beyond 20 will only affect power if the fraction of missing is very high. Each data set compiled different estimates for the parameters of interest. The statistical analysis is a regression analysis, and it consists of analyzing every imputed data set, which will all be complete. All the missing values now have new estimated values; the analysis will be of a complete data. The tiresome process of analyzing up to twenty newly created imputed regression analysis are made easier by software packages with built in routines. The analysis phase is the easiest part of the multiple imputation analysis.

The pooling phase combines all the estimates into each parameter are possibly unbiased if the data is MAR or MCAR. Instead of using a single imputed data to estimate parameters, multiple imputation analysis combines all the newly imputed estimates into single point estimates. Rubin (1987) outlined formulas for pooling parameter estimates and standard errors. The pooled parameter estimates are the mean value of all the imputed m estimated values

$$\bar{\beta} = \frac{1}{m}\sum_{t=1}^{m} \hat{\beta}_t, \tag{2.13}$$

where $\hat{\beta}_t$ is the parameter estimate from data set t and $\bar{\beta}$ is the pooled estimate. Combing the total standard errors uses a within and between source of variation. The within variation is the mean value of the squared standard errors and compile the fluctuation that would have resulted if there was no missing data. The within imputation variance is the average of the m sampling variances

$$V_w = \frac{1}{m}\sum_{t=1}^{m} SE_t^2, \tag{2.14}$$

where V_w denotes the within imputation variance, and SE_t^2 is the squared standard error from data set t. The between variation accounts for the estimates across all the imputed estimates. The between imputation variance quantifies the variability of a parameter estimate across the m data sets and has the following equation:

$$V_B = \frac{1}{m-1}\sum_{t=1}^{m}(-_t - \bar{\beta})^2 \tag{2.15}$$

where V_B denotes the between imputation variance, $\hat{\beta}_t$ is the parameter estimate from data set t and $\bar{\beta}$ is the average point estimate from Equation 2.13. The total sampling variance combines the within imputation variance with the between imputation variance, as follows

$$V_T = V_w + V_B + \frac{V_B}{m}. \tag{2.16}$$

The total sampling variance is the sum of the within and between imputation components plus the calculation of $\frac{V_B}{m}$, which is the average parameter estimate. Together the within and between imputation variance account for the total error due to the missing data.

Researchers are nervous about using multiple imputation procedures because the procedure “creates data” while others say it “makes up data.” However, multiple imputation creates an average of parameter estimates that account for the uncertainty of missing data. The purpose of multiple imputation is to have proper inferential methods, use evidence we have of what we are missing to fix our hypothesis tests.

Single-level multivariate multiple imputation rationale conditions the predictors to preserve the relationship among the outcome in the imputed data. However, practical problems often arise when imputing multivariate missing data. For example, the variables are often diverse types (e.g., binary, ordered, continuous) making a convenient model like multivariate normal theoretically inappropriate. The relationship between the outcome and the predictors can be complex and nonlinear. Over time, multiple imputation has developed many different methods for the diversity of missing data problems; however, not one single method works best for all situations. The two main strategies for imputing single-level multivariate data are joint modeling (JM) imputation (Rubin & Schafer, 1990; Schafer, 1997) and fully conditional specification (FCS) imputation (van Buuren et al., 2006).

The single-level JM assumes incomplete variables follow a multivariate normal distribution. Van Buuren (2018) indicates that if a joint model is specified, it is nearly always the multivariate normal model. Commonly using a multivariate normal model for all incomplete variables JM draws missing values simultaneously. The missing values are imputed using a joint model (e.g., the multivariate normal model). Meanwhile, single-level FCS (also referred to as imputation by chained equations) imputes multivariate missing data on a variable-by-variable basis drawing values from a series of univariate conditional distributions. This requires specification of a separate imputation model for each incomplete variable. Missing values are imputed one variable at a time until all the missing values in the variable are filled. This complete data predictor is used in the next imputation model continuing until the algorithm cycles iteratively through all the incomplete variables. Hughes et al. (2014) concluded that FCS and JM imputation are equivalent in single level data sets with multivariate normal variables.

Missing values in multilevel data adds to the complexity of using multiple imputation. The imputation model must account for random effects to correctly manage the clustering in the data. Multilevel missing imputation procedures use imputation models based on the linear mixed-effects model. The JM and FCS have extended approaches for multilevel imputation. The JM (Schafer, 2001; Schafer & Yucel, 2002; Yucel, 2008) specifies a single model for all incomplete variables in data. The FCS (van Buuren, 2011) iterates univariate multilevel imputation over the variables. Researchers have compared the JM and FCS multilevel approaches and found that both the JM and FCS imputation are appropriate for random intercept analyses, finding unbiased estimates for balanced data and normally distributed variables when the missing data mechanism is MAR or MCAR (Enders et al., 2016; Kunkel & Kaizar, 2017; Mistler & Enders, 2017). Enders et al. (2016) found the JM method superior for analyses that focus on within and between cluster associations while FCS provided dramatic improvement over the JM in random slope models. Van Buuren (2018) stresses that there is not one “super” method that will address all longitudinal missing data issues. Several extensions of the standard JM and FCS approaches for imputing diverse types of missing longitudinal/cluster data issues have been proposed in the literature over recent years (Enders et al., 2018; Goldstein et al., 2009; Nevalainen et al., 2009; Quartagno & Carpenter, 2016; Resche-Rigon & White 2018; van Buuren 2011;). For further research on the multilevel extensions, Huque et al. (2018) provide an overview of twelve different MI techniques that include the standard FCS and JM methods plus eight FCS and two JM extensions. They conducted a simulation study to compare imputed incomplete longitudinal covariates results for all the methods and concluded that the FCS and JM standard methods performed well.

Missing not at Random Missing Data Models

There are two types of MNAR missing data techniques researchers typically choose, the selection model and the pattern mixture model. Both methods use a joint distribution to describe the data and the probability of missingness. Although, the two methods attempt to do so in vastly different ways. Both models include an additional component into the estimation process to decrease or eliminate bias that results from the MAR methods.

Selection Models

The classic selection model (Heckman, 1976) was proposed for MNAR data as a method for correcting bias in a regression model. It is a two-part model that combines the substantive analysis with an additional regression equation that models response probabilities. For the selection model, the probability of missingness represented by P the probability distribution, $\underline{\underline{\boldsymbol{X}}}$

represents the sample data, and M is the corresponding missing data indicator

$$P(\boldsymbol{X}_{it}, M) = P(M|\boldsymbol{X}_{it}) * P(\boldsymbol{X}_{it}). \tag{2.17}$$

The two-parts of the selection model use the joint distribution into the product of $P(M|\boldsymbol{X}_{it})$ as the conditional distribution of missingness, given $\boldsymbol{X}_{it}$ (sample data), and $P(\boldsymbol{X}_{it})$ is the marginal distribution of the data. $P(\boldsymbol{X}_{it})$ is the part of the model that would have been estimated with no missing data. Here is a look at this regression model

$$y_{it} = \beta\boldsymbol{X}_{it} + \varepsilon_1 \tag{2.18}$$

where y_{it} denotes the dependent variable, $\boldsymbol{X}_{it}$ denotes the independent variables, β denotes the parameters to be estimated and ε_1 is an error term that is normally distributed with a mean of zero and a standard deviation of σ. The conditional distribution defines the probability that a participant with a particular value of X has missing data. This is the second part of the selection model $P(M|\boldsymbol{X}_{it})$ that predicts response probabilities through the regression equation

$$M^* = \beta_0 + \ \beta_1 X + \varepsilon_2. \tag{2.19}$$

Where M^* is not the binary missing indicator but is an individual's latent propensity for missing data, β_0 and β_1 denote the regression intercept and slope, and ε_2 is a normally distributed residual term with a mean of zero and a standard deviation of one. The conditional probability distribution in the regression model describes the probability that a participant with a score has a missing value. Correlated residuals link the regression model with missing data correcting for bias in the substantive model. The equation for the correlated residuals from the error terms in Equations 2.18 and 2.19 above is

$$corr(\varepsilon_1, \varepsilon_2) = \rho, \tag{2.20}$$

where ρ represents the correlation between the two error terms.

Diggle and Kenward (1994) adapted the selection model for longitudinal analyses for data with monotone missingness. The method combines growth curve analysis with regression equations that predict response probabilities. Their model assumes the missingness mechanism is MNAR combining the LMM with a logistic regression for the dropout process. The margin model for $\boldsymbol{Y}_{it}$ is combined with a model for the dropout process, conditional upon if there is a measurement by the participant. Here is the model

$$f(\boldsymbol{y}_{\boldsymbol{it}}, D_{it} | \beta, \varphi) = \int f((\boldsymbol{y}_{\boldsymbol{it}}, D_{it} | \beta, \varphi) dy_i^m = \int f(\boldsymbol{y}_{\boldsymbol{it}} | \beta) f(D_{it} | \boldsymbol{y}_{\boldsymbol{it}}, \varphi) dy_i^m \qquad (2.21)$$

where i refers to the participant and $i = 1, \ldots, N$, where N is equal to the total number of participants. The outcome $\boldsymbol{Y}_{it}$ is measured at time point t for each participant and is allowed to differ, therefore, $t = 1, \ldots, n_i$, resulting in a vector of observed outcomes. The term D_{it} is the occasion where dropout occurs and is the second part of the model implementing a logistic regression for the binary missing data indicators that describe the likelihood of dropout at each wave of data collection. The model assumes the measured variables are multivariate normal. The logistic dropout model is

$$logit[P(D_i = t \mid D_i \geq t, \ \boldsymbol{y}_{\boldsymbol{it}}, \ \varphi)] = \varphi_0 + \varphi_1 \boldsymbol{y}_{\boldsymbol{it}} + \varphi_2 \boldsymbol{y}_{\boldsymbol{i,t-1}}, \qquad (2.22)$$

where vector $\boldsymbol{Y}_{it}$ refers to collected longitudinal responses for each participant. If dropout occurs, $\boldsymbol{Y}_{it}$ is partially observed. The drop out term is D_i and denotes the occasion t at which subject i drop out occurs. The conditional probability $P(D_i = t \mid D_i \geq t, \ \boldsymbol{y}_{\boldsymbol{it}}, \ \varphi)$ is used to calculate the probability of dropout at each measurement occasion. Others have used this approach to analyze longitudinal data that involves missing mechanisms assumed to be MNAR (Little, 1995; Molenberghs & Kenward, 2007). For non-monotone missing data, Ibrahim et al. (2001) proposed a method for estimating parameters in the generalized linear mixed-effects model using a selection model with non-ignorable missing response data. The random coefficients selection model is another model for longitudinal data analysis (Little, 1995; Shih et al., 1994). This model uses individual growth curves to predict the probability of missing data.

Pattern Mixture Models

The pattern mixture model (Glynn et al., 1986; Little, 1993, 1994) creates subgroups of cases that share a similar missing data pattern and estimates the substantive analysis model from each pattern. For pattern mixture models the probability of missingness represents P the probability distribution, X represents the sample data, and M is the corresponding missing data indicator

$$P(\boldsymbol{X}_{it}, M) = P(\boldsymbol{X}_{it}|M) * P(M). \tag{2.23}$$

The conditional distribution $P(\boldsymbol{X}_{it}|M)$ for the sample data given a particular value of M, and $P(M)$ is the marginal distribution of missingness. The conditional distribution is the probability of obtaining different X values within a subgroup that share the same missing pattern. The marginal distribution describes the different missing data patterns. The cases are stratified into subgroups, which provides parameter estimates for each. Then a computed weighted average compiles the stratified specific estimates into a single set of estimates. The pattern mixture model extends to longitudinal analysis by estimating the growth model separately for each missing data pattern by averaging their regression coefficients into a single estimate. The linear mixed model from Equation (2.1) adds in all dropout patterns

$$\boldsymbol{Y_{it}} = \beta_{00} + \beta_{10}X_{it} + \beta_{01}D_i + \beta_{11}D_iX_{it} + \mu_{0i} + \mu_{1i} + \varepsilon_{it}. \tag{2.24}$$

The new term is D_i, which is a factor that puts the participants into missing data pattern groups. The following interpretation for the rest of the regression coefficients are: β_{00} is the baseline for completers, β_{10} is the growth rate for completers, β_{01} is the baseline mean difference between the completers and the dropouts, and β_{11} is the growth rate difference between the two patterns.

Note this is the simplest case of two groups; pattern mixture models can have many dropout patterns.

The literature for pattern mixture models on MNAR longitudinal data focuses on maximum likelihood methods of a mixed-effects model with normally distributed outcomes. Typical research assumes monotone missing data patterns where participants miss a measurement occasion and drop out for the rest of the study. The researcher can then use the point where participants drop out as natural forming patterns and compare them with the completers for the pattern mixture model (Hogan & Laird, 1997; Little, 1995; Molenberghs et al., 1998; Thijs et al., 2002). Extensions of pattern mixture models allow random effects to be included in a pattern mixture model. Random coefficient pattern mixture models (Demirtas & Schafer, 2003; Fitzmaurice et al., 2001; Hedeker & Gibbons, 1997; Little, 1995) divide the subjects into groups as subject level covariates based on their missing data pattern and examine the effect of the different patterns on the outcome of interest. Using these models, we try to capture an underlying process that the drop out or missing data are related to the outcomes.

Pattern mixture models have been adapted to the longitudinal missing data scenario of non-monotone missingness. Latent class pattern mixture models (H. Lin et al., 2004; Roy, 2007) explore intermittent missing data by forming patterns based on latent classes. Roy (2007) describes latent class models that could be used for characterizing missing data patterns in longitudinal studies with regularly spaced observation times, where there are high percentages of intermittent missing data. H. Lin et al. (2004) analyzed the missingness process in the form of latent classes that were conditionally independent of the longitudinal outcomes that they named the latent pattern mixture model. In this work, the data contained intermittent and monotone missing data where mixture patterns are formed from latent classes that link the longitudinal response and the missing process. Here is a look at the model defined participant class membership to a latent class

$$\pi_{ik} = P(C_{ik} = 1) = \frac{\exp(X_i^{\mathrm{T}}\eta_k)}{\sum_{j=1}^{k}\exp(X_i^{\mathrm{T}}\eta_j)}. \tag{2.25}$$

In this equation, i refers to the participant and $(i = 1, \ldots, n)$ and K latent classes labeled $(k = 1, \ldots, K)$. Let $c_i = (c_{i1}, \ldots, c_{ik})^{\mathrm{T}}$ be the multinomial distributed class membership vector for participant i with $c_{ik} = 1$ if subject i belongs to latent class k and 0 otherwise. The probability that subject i belongs to latent class k is π_{ik}, which is modeled by a logit model including covariate vector $X_i = (X_{i1}, \ldots, X_{im})^{\mathrm{T}}$ and associated class specific coefficient vector η_k with $\eta_1 = 0$. The model will determine posterior probabilities for each participant into a latent class. The latent pattern mixture model creates arbitrary patterns of missing data represented by the visit process of the participants and avoids the need to specify the missing patterns a priori. Implementing the latent pattern mixture model in the H. Lin et al. (2004) study suggested the presence of four latent classes linking the participant visit patterns to the outcomes.

Intensive Longitudinal Data

Rationale for Intensive Longitudinal Data (ILD) and Ecological Momentary Assessment (EMA) is to study participants in their daily real-world setting that avoids recall bias and enables the analysis of dynamic processes over time. The methods include the collection of short self-reports repeatedly during daily activities, thus allowing events, behaviors, sensations, thoughts, feelings, emotions, mood, symptoms, and actions to be monitored in the individuals' natural settings (Messiah et al., 2011). ILD studies often examine the relationships between momentary contextual variables and a behavior of interest. Examples of types of EMA studies include work activity and satisfaction, pain levels, relationships, psychotherapy, drug and alcohol use, physical activities, and psychological stress to name just a few. ILD aims to capturing current states and events of the participants as they are lived hour-to-hour and day-to-day. For example, if one is interested in the experience of substance use, assessing behaviors will be better in the context of how they naturally occur instead of a research clinic. These types of studies offer the benefit of studying within participant change over time that is absent in typical survey research. EMA studies use the temporal resolution afforded by multiple measures to focus on the within-subject changes in behavior and experience over time and across contexts, addressing how symptoms vary over time or how situational antecedents influence behavior (Shiffman et al., 2008).

Recent technological developments have made the collection of ILD more convenient for participants and researchers. ILD studies assess particular events in subjects' lives or assess subjects at periodic intervals, often by random time sampling, using technologies ranging from written diaries and telephones to electronic diaries and physiological sensors (Shiffman et al., 2008). Researchers give participants a small electronic device or have them download an APP to their own private smart phones to collect responses. The device or APP prompts participants at various times throughout the day recording the responses with time stamps of compliance as participants complete each assessment. The method allows for a more accurate measurement of outcomes as they occur in a natural setting. The participants in the study are instructed to continue their daily lives, which means they may be in situations that make it inconvenient for them to respond to the prompting device. For example, in ILD collected for college students, a participant could be with a group of friends, in class, or studying and not want to answer a prompt. In other instances, they could turn off their device at specific times to avoid being bothered (i.e., driving). Sokolovsky et al. (2014) indicated that individuals may not respond to all prompts or cues to report experiences or may otherwise systematically avoid reporting; both instances may introduce important biases into data collection.

ILD methods often collect a large number of observations per participant where prompts occur at random times throughout each day. The random nature of the prompting mechanism inevitably leads to some participant having nonresponse on the prompts. The missingness is usually prompt-wise because the data are rarely partial for a given prompt: Participants typically either ignore the signal entirely (causing all items to be missing for that questionnaire) or they respond to all items (McLean et al., 2017; Silvia et al., 2013). As participants miss an assessment, no information is collected at the time and challenges researchers to predict the probability of missingness. Missing assessments have the potential to bias the obtained sample of behavior and experience, especially if the missing data mechanism is non-random (Shiffman et al., 2008). Many stress the importance of missing data techniques in ILD studies (Stone et al., 2007; Walls & Schafer, 2006).

The nature of ILD has created new missing data challenges for researchers as the pattern of missingness is non-monotone for the several assessments among study participants. Intermittent missing data in ILD studies may include many unique patterns among study participants making pattern mixture models hard to implement. The common approach to handling missing data in ILD studies is to make mention of the missing data but not to present missing data models for the missingness. This assumes the missingness is MCAR and allows the statistical software to handle the missing data, which typically implements list-wise deletion (Peugh & Enders, 2004). Improper methods such as deletion of records with missing variables, mean substitution, or simple imputation methods to account for missing data cause bias and decreased efficiency (Albert & Follmann, 2009; Enders, 2010). Intermittent missingness can be an issue in ILD studies as participants with higher proportions of non-response may behave differently in terms of the outcome compared to those with lower proportions. The missingness may impact the within-subject variation of the outcome on the participants. For example, participants that respond more often might have more consistent variation in outcomes than those that respond less often. As participants miss prompts it alters the time intervals between data points, which is an important aspect in the studying if the behavior at a given time point is associated with behaviors at subsequent time points. Intermittent missing data within ILD studies can potentially bias parameter estimates and has unique properties that make longitudinal missing data models challenging to implement. Two missing data methods have emerged with shared parameter models that provide information about the missingness process and the outcome that were tailor made for ILD studies. The ensuing sections will introduce patterns of missing data, the mixed-effect location scale model, multiple imputation model, and two shared parameter models that constitute the recent literature for analyzing missing ILD.

Patterns of Missing Data

Missing data patterns refer to the arrangement of observed and missing values in a data set. The two patterns of missing data that will need to be described for this dissertation are monotone and non-monotone missing patterns. Data follow a monotone missing pattern when a subject misses an observation occasion and never is observed again. Monotone missing data is often termed dropout or attrition. Participant dropout commonly occurs in longitudinal studies as each study is guaranteed to lose participants. Data follow a non-monotone missing pattern if a subject has observed values after a missing value occurs. Data that follow this pattern are termed intermittent missing. Intermittent missingness is common in ILD studies as participants skip a prompt and then respond to future prompts.

Mixed-Effect Location Scale Model

Hedeker et al. (2008) described the mixed-effect location scale model for the context of EMA intensively measured longitudinal studies. The work extends the LMM from Equation 2.1 by adding a subject level random scale effect to the within-subject variance specification. This allows the within-subject variance to vary at the participant level. The reason to include both location and scale random effects is to allow for participant heterogeneity in both the mean and within participant variability of the outcome that cannot be fully explained by covariates. This relaxes the homogenous error variance assumption adopted by most statistical methods. The variance of $\log(\varepsilon_{it})$ accounts for individuals' distinct patterns for the outcome. Here is the LMM from Equation 2.1 where u_{0i} is the random intercept location parameter for the participants:

$$y_{it} = \beta_0 + \boldsymbol{X_{it}^T\beta} + u_{0i} + e_{it}. \tag{2.26}$$

Here is the model for the scale effects extension in the within-subject variance model:

$$\log\left(\sigma_{e_{it}}^2\right) = \boldsymbol{Q_{it}^T\tau} + u_{2i}, \tag{2.27}$$

where the log function ensures that the estimated error variance is strictly positive. ε_{it} reflects the fluctuations on the outcome measured for participant i, thus the smaller the variance the more stable participant i is on the outcome. The term $\boldsymbol{Q_{it}}$ is the within subject variance model and usually contain a subset of the variables in $\boldsymbol{X_{it}}$ allowing time dependent covariates. The term τ is the fixed effect coefficient vector and indicate the effect of the within subject coefficient on the log-variance of the outcome. The random subject scale intercept is u_{2i} indicating the effect of participant *i* on his/her within subject variability of the repeated measurements. The random effects $\{u_{0i}, u_{2i}\}$ are assumed to follow a bivariate normal distribution with mean 0 and a covariance structure

$$\boldsymbol{G} = \begin{bmatrix} \sigma^2_{\mu_{0i}} & \sigma^2_{\mu_{0i,2i}} \\ \sigma^2_{\mu_{0i,2i}} & \sigma^2_{\mu_{2i}} \end{bmatrix}. \tag{2.28}$$

Conditional on $\{u_{0i}, u_{2i}\}$, the outcome measurements Y_{it} are *i.i.d* normal. The mixed-effect location scale model is estimated using maximum likelihood estimation and significance tests for the fixed effects model parameters are typically done using the Wald tests.

Intensive Longitudinal Data with Multiple Imputation

Multiple imputation (MI) works the same for missing data in ILD studies as it does in longitudinal studies described in multiple imputation section. Ji et al. (2018) presented a new way to combine both MAR longitudinal missing data methods of FIML (full information maximum likelihood section) and of MI in the context of ILD missing data. They termed the approach partial MI, which performs MI on the missing covariates while missingness for the dependent variable implement FIML estimation. The other missing data models in the study included two full MI approaches that impute values for both the missing covariates and dependent variables. The hypothesized model is vector autoregressive (VAR) that predicts the dependent variable at the current time point *t* by the dependent variable at the immediately preceding time point *t-1*, often referred to as a time lag. Here is a look at the typical VAR model with a lag of one equation:

$$Y_t = \beta_0 + \beta_1 Y_{t-1} + e_t \tag{2.29}$$

where the dependent variable at the current time point is Y_t. The intercept is β_0, and β_1 is the coefficient for the dependent variable of the lag for the preceding time point and e_t is the error term. Note that VAR models can have more than one time lag in the model. All predictors in the model have an equation explaining its development based on their own lagged values, the lagged values of the others in the model predictors, and an error term.

The two MI approaches impute data in different ways. The first MI approach is the MICE approach described by van Buuren and Groothuis-Oudshoorn (2011), which is the FCS multilevel approach discussed more thoroughly in the longitudinal modern missing data methods described earlier in the multiple imputation section. To review, MI approach is implemented via chained equations where imputations are drawn by iterating over the conditional densities on a variable-by-variable basis by means of the Markov chain Monte Carlo (MCMC) technique. MICE allows flexibility among the variables as it imputes depending on the distributional characteristics of the variables to be imputed (e.g., normal continuous data, ordinal, nominal). The second MI approach implemented Ameilia II multiple imputation program specializes in handling missingness in time-series data (Honaker et al., 2011). This MI has a built-in feature that is set up for variables in the lagged VAR models to be imputed. Amelia II performs imputations by assuming the variables are multivariate normally distributed with a mean vector, μ, and covariance matrix, Σ.

A simulation was conducted to compare the performance of both full MI methods, partial MI approach, and list-wise deletion under different missingness conditions MCAR, MAR, and MNAR. The results of the study have provided context to applying longitudinal MAR missing data models MI and FIML in ILD studies (Ji et al., 2018). The performance of the MI approaches, Amelia II, MICE, and the partial MI were better than using list-wise deletion regarding smaller biases in the point estimates, especially for time-dependent covariates. The researchers included multiple time points of 15 and 75 but doing so did not provide any improvement in accuracy of the point estimates. Every MI approach improved the accuracy of the standard error estimates over list wise deletion. However, under MNAR missing condition, the estimates from the full MI method Amelia II had higher biases than other missing data models, including list-wise deletion. The partial MI approach that performs MI on the missing covariates while missingness in the dependent variables apply FIML emerged as the top option compared to the full MI approaches for the covariates, time-dependent covariates and dependent variables including when the data was MNAR. The partial MI method had results of better accuracy, precision, and coverage compared to both full MI approaches MICE and Ameilia II. The results of Ji et al. (2018) provide valuable information regarding a new technique that involves a combination of MI and full information likelihood missing data models for EMA studies.

Shared Parameter Model with Item Response Theory

The initial development of Item Response Theory (IRT) took place when Rasch (1960) developed the first model for analyzing categorical data, known as the Rasch model. Lord and Novik (1968) followed it with the theory of latent trait estimation, changing the method of data analysis in testing. IRT is a collection of statistical and psychometric methods used to model test takers' item responses (Yen & Fitzpatrick, 2006). IRT defines a scale for the underlying latent variable that is being measured by the test items. The model specifies how both trait level and item properties are related to a person's item responses. The following equation represents the two-parameter logistic IRT model:

$$P(\mu_i = 1|\theta_j, \alpha_i, \beta_i) = \frac{e^{\alpha_i(\theta_j - \beta_i)}}{1+e^{\alpha_i(\theta_j - \beta_i)}}, \qquad (2.30)$$

where i corresponds to the item on the test and j corresponds to the participant taking the test. The latent trait θ_j is the test takers "ability" for participant j, which is created using observed responses to the items i on a test. In a two-parameter logistic IRT model, α_i is the discrimination parameter for each item i and is seen as the slope that discriminates between the test takers who know the right answer and the population of test takers who do not demonstrate that knowledge.

The difficulty parameter β_i for each item i determines the manner of which the item behaves along the ability scale. The difficulty parameter is on the same scale as the test takers' ability (θ) and are estimated separately. The difference between difficulty and ability parameters provides information about the probability or log odds of a correct response for each item. For example, if $\theta_j > \beta_i$ it means that the examinee's ability level is greater than the item difficulty making the item easier for them, conversely, if $\theta_j < \beta_i$ means the item is difficult for the examinee. In the next few paragraphs, there will be a demonstration on how IRT is a missing data model for EMA by letting the prompting time-bins (how researchers collect data) represent items in an IRT model.

Cursio et al. (2019) modeled the intermittent missing prompts as a continuous latent trait using IRT called the Latent Trait Shared Parameter Mixed Model (LTSPMM). The IRT model represents the latent trait of "responsiveness" and corresponds to how each participant responds to the prompting device (electronic device or APP). The IRT response mechanism for EMA data uses the time-bins as the items. For example, if the ILD study design prompts the participants on forty-two occasions then there will be a possible 42 time-bins that correspond with the days and times available for the participants to respond. The dichotomous outcome variable in the IRT model is defined as responding R_{it}, where R_{it} has a value of 1 for participant i if the participant responded to the prompt in time-bin t and has a value of 0 if a prompt was not answered. Cursio et al. (2019) applied a one-parameter (1PL) and two-parameter (2PL) logistic IRT model as the probability of response to the prompting device.

The IRT model uses a latent trait (θ_i) that represents participants "responsiveness" and models jointly with the LMM for longitudinal outcomes. This is the longitudinal mixed model used by Cursio et al. (2019):

$$y_{it} = \beta_0 + \boldsymbol{X_{it}^T\beta} + u_{0i} + e_{it}. \tag{2.31}$$

The outcome is represented by y_{it} where i represents the number of subjects and t represents the number of repeated observations within each subject. The matrix $\boldsymbol{X_{it}}$ contains the subject-level and time-dependent covariates including a time predictor in the model where the first column is a vector of ones for the intercept terms. The random intercept term u_{0i} are assumed to be normally distributed $u_{0i}{\sim}N(0, \sigma_\mu^2)$ and accounts for the intra-subject correlation due to the multiple responses from each subject. The error term e_{it} is assumed to be normally distributed $e_{it}{\sim}N(0, \sigma_e^2)$. The one-parameter (1PL) and two-parameter (2PL) logistic IRT models have the following logistic form modeling the probability of responding to a prompt:

$$P(R_{it} = 1|\theta_i) = \frac{1}{1+\exp[-a_t(\theta_i - b_t)]} \tag{2.32}$$

where R_{it} represents responding to time-bin t from subject i. The $P(R_{it})$ is the probability that person i responds to time-bin t, which is similar probability of answering the correct answer to item i on a test from Equation 2.30. The latent trait responsiveness is represented by θ_i and b_t is the difficulty parameter corresponding to each time-bin t. The discrimination a_t parameter provides the slope for each time-bin t. The 1PL sets a fixed slope a across all time-bins while the 2PL allows for unique slopes a_t for each time-bin accounting for more information in the model. Allowing the slopes to vary is the only difference between the 2PL and 1PL models. The number of participants ranges from $1\ to\ N$, and the number of time-bins t ranges from 1 to m_i allowing for a different number of time-bins for each subject. The model represents the log odds of responding to the prompting time-bin and displays the responsiveness of each participant. Slightly changing the form of the 1PL and 2PL models

$$P(R_{it} = 1|\theta_i) = \frac{1}{1+\exp[-(c_t + a_t\theta_i)]}, \tag{2.33}$$

where the time-bin intercept parameter is $c_t = -a_t b_t$. The term $c_t$in the 2PL model, the discrimination parameters a_t vary by time-bin t while in the 1PL model the discrimination parameters remain constant. Thus, writing the 1PL and 2PL models in terms of the log odds of responding

$$log\left[\frac{P(R_{it}=1|\theta_i)}{1- P(R_{it}=1|\theta_i)}\right] = c_t + a_t\theta_i. \qquad (2.34)$$

The item difficulty parameters (b_t) and discrimination parameters (a_i) provide information about the time-bins. The difficulty parameters give information regarding the difficulty the participants have responding at a particular time-bin. In this context, the prompting time-bin with the most responses have the lowest difficulty and the time-bin with the most missing responses have the highest difficulty. As mentioned above, the difference between the 1PL and 2PL logistic IRT models is that the 2PL estimates discrimination parameters (a_i) for each time-bin. The discrimination parameter describes how the probabilities change between the latent traits of responsiveness for the participants. In this context, a prompting time-bin may clearly discriminate between high and low responders. The 1PL and 2PL IRT models both estimate the time-bin difficulty parameters (b_t) for all the time-bins in the study. In the 2PL LTSPMM, a discrimination parameter is estimated for each time-bin. Estimating the difficulty and discrimination parameters for each prompt consumes several degrees of freedom thus causing the LTSPMM to have issues with convergence and computation processing time. Convergence and slow processing speeds are a drawback of using the LTSPMM missing data method (Cursio et al., 2019).

The LTSPMM portion of the model is an expansion of the random intercept into the equation written as

$$u_{0i} = \gamma\theta_i + \eta_{0i}. \qquad (2.35)$$

The term γ represents the effect of the latent trait θ_i of responsiveness on the outcome y_{it} for participant i, which is normally distributed $\theta_i \sim N(0, \sigma_\theta^2)$. The value of θ_i behaves as a random effect and influences the log odds of responding to the prompt. The random-intercept term will allow for participants to have unique intercepts and is written in the form of η_{0i}, which is normally distributed as $\eta_{0i} \sim N(0, \sigma_\eta^2)$. The expansion adds into the LMM of Equation 2.31 and results in the following model:

$$y_{it} = \beta_0 + \boldsymbol{X}_{it}^{\boldsymbol{T}}\boldsymbol{\beta} + (\gamma\theta_i + \eta_{0i}) + e_{it}. \tag{2.36}$$

The latent trait θ_i is linked for the full LTSPMM by the 1PL or 2PL IRT models. In this approach, the generalized linear mixed model is used to estimate a latent trait for responsiveness and a separate latent trait for the outcome. The outcome $\boldsymbol{y}_{it}$ is influenced by the latent trait for response θ_i, which is a shared parameter in each sub model. As a reminder, here is the logistic form modeling the probability of responding $P(R_{it})$ to a prompt from Equation 2.34:

$$log\left[\frac{P(R_{it}=1|\theta_i)}{1- P(R_{it}=1|\theta_i)}\right] = c_t + a_t\theta_i. \tag{2.37}$$

In the simulations using the LTSPMMs, Cursio et al. (2019) found that the latent trait of participant “responsiveness” coefficient was significant for both the 1PL and 2PL models. Thus, demonstrating that the ability to respond to the prompting device had an influence on model outcomes. In the simulation, both LTSPMMs outperformed a full mixed-effect location random effects model using list-wise deleted model in terms of bias and coverage rates for the true model of coefficients gender and negative mood regulation under the MNAR missingness conditions as gender was set to correlate with missing. List-wise deletion underperformed when the missing data had MNAR properties for the estimated regression coefficient gender supporting the use of the LTSPMM. The 2PL outperformed the 1PL LTSPMM in terms of bias and standard error of the estimated regression coefficient of the latent trait θ_i. A likelihood ratio test comparing the

1PL and 2PL LTSPMM was significant indicating that the additional discrimination parameters in the 2PL model did improve the overall fit. However, the 2PL LTSPMM did take much longer to fit and converge.

Model Estimation

The model estimation for the LTSPMM uses a maximum marginal likelihood estimation described by Bock and Aitkin (1981). The full data likelihood needs to be averaged over the two random effects represented by the random intercept η_{0i} and the latent trait θ_i (Liu, 2008; Liu & Hedeker, 2006). The LTSPMM can be written as:

$$L(\boldsymbol{Y}|\mu,\theta,R) = \prod_{i=1}^{n}\prod_{t=1}^{n_i} f_1(\boldsymbol{y_{it}}|\, x_{it}, u_{0i}, \sigma_\mu, \sigma_e^2). \tag{2.38}$$

with error terms that are assumed to be normally distributed with standard deviation σ_e

$$e_{ij} \sim N(0,\sigma_e^2). \tag{2.39}$$

The u_{0i} term includes the effect of the latent trait θ_i and a random intercept η_{0i}. The term u_{0i} is defined as the sum of $\gamma\theta_i$ and $\sigma_\mu\eta_{0i}$.

$$p(\mu|\theta,R) = \prod_{i=1}^{n}\prod_{t=1}^{n_i} f_2(R_{it}|\, \theta_i, \eta_i) \tag{2.40}$$

$$\theta_i \sim N(0,1) \text{ and } \eta_{0i} \sim N(0,1). \tag{2.41}$$

Combining the random effects into Equation (2.9) results in the following expression for the LTSPMM:

$$L\big(\boldsymbol{Y},R\big|\theta,\eta;\ \beta,\sigma_e^2\big) = \prod_{i=1}^{n}\prod_{t=1}^{n_i} f_{1,2}(\boldsymbol{y_{it}}|\, x_{it},\theta_i,\eta_{0i},\sigma_e^2) f_2(R_{it}|\theta_i,\eta_i). \tag{2.42}$$

Since the probability density function of η is a standard normal with mean 0 and standard deviation 1 then:

$$p(\eta_{0i}|\sigma_\mu) = (2\pi\sigma_\mu)^{1/2} exp\left[\frac{1}{\sigma_\mu^2}\eta_{0i}^2\right]. \tag{2.43}$$

The maximum likelihood estimation in the distribution is obtained by integrating over the distributions of θ and η, the marginal distribution for the LTSPMM can be written as:

$$\mathrm{h_i}(y_\mathrm{i}) = \int_\theta \int_\eta f(\boldsymbol{y_i}|\theta,\eta,\beta,\sigma_e^2)\ g(\theta,\eta;\ \Sigma_{\theta,\eta})\ d\theta_i\, d\eta_i \tag{2.44}$$

with the probability density $f(\boldsymbol{y_i}|\theta,\eta,\beta,\sigma_e^2)$ equal to

$$f(\boldsymbol{y_{it}}|\theta,\eta,\beta,\sigma_e^2) = (2\pi)^{-n1/2}|\sigma_e^2\boldsymbol{I_{n_i}}|\exp\left[-\frac{1}{2}(\boldsymbol{y_{it}} - \boldsymbol{X_{it}^T\beta} - \gamma\theta_i - \sigma_u\eta_i)'(\sigma_e^2 I_{n_i})^{-1}(\boldsymbol{y_{it}} - \boldsymbol{X_{it}^T\beta} - \gamma\theta_i - \sigma_u\eta_i)\right] \tag{2.45}$$

and the density for the random effects is a normal distribution written as:

$$g(\zeta;\ \Sigma_\zeta\ ) = (2\pi)^{-1}|\ \Sigma_\zeta\ |^{-1/2}\exp\left[-\frac{1}{2}\zeta'\,\Sigma_{\boldsymbol{\zeta}}^{-\mathbf{1}}\,\boldsymbol{\zeta}\right]. \tag{2.46}$$

The vector ζcontains the two standardized random effects θ and η and $g(\,)$ is a multivariate standard normal pdf. Summing the marginal log-likelihoods over the sample and maximizing the function β, σ_e^2, θ, and η gives the values. To denote the posterior density p_i, likelihood function $f_{1,2}$, prior density g, and the marginal log-likelihood h_i. The maximum likelihood solution for the fixed effects coefficient vector β is derived as:

$$\log L = \sum_{i=0}^{N}\log h_i\,(\boldsymbol{y_i})$$

$$= \sum_{i=1}^{N}\log\left[\int_\theta \int_\eta f(\boldsymbol{y_i}|\theta,\eta;\ \beta,\sigma_e^2)\ g(\theta,\eta;\ \Sigma_{\theta,\eta})\, d\theta_i\, d\eta_i\right]$$

$$p_i = p(\theta_i) \times p(\eta_i)$$

$$= \prod_{i=1}^{n}\prod_{t=1}^{n_i}(\boldsymbol{y_{it}}|\,x_{it},\theta_i,\eta_{0i},\sigma_e^2)(R_{it}|\theta_i,\eta_i)\cdot\frac{(2\pi)-1|\,\Sigma_\zeta\ |-1/2\ \exp\left[-\frac{1}{2}\zeta'\,\Sigma_{\boldsymbol{\zeta}}^{-\mathbf{1}}\,\boldsymbol{\zeta}\right]}{\int_\theta \int_\eta f(y_i|\theta,\eta,\beta,\sigma_e^2)\ g(\theta,\eta;\Sigma_{\theta,\eta})\ d\theta_i\, d\eta_i}$$

Taking the derivative of both sides with respect to β results in:

$$\frac{\partial logL}{\partial\beta} = \sum_{i=1}^{N}\frac{\partial\log h_i}{\partial\beta}$$

$$= \sum_{i=1}^{N} \frac{1}{h_i} \frac{\partial\left[\int_\theta \int_\eta f(\boldsymbol{y_i}|\theta,\eta;\beta,\sigma_e^2)\, g(\theta,\eta;\Sigma_{\theta,\eta})\, d\theta_i\, d\eta_i\right]}{\partial\beta}$$

$$= \sum_{i=1}^{N} \frac{1}{h_i} \int_\theta \int_\eta \frac{\partial f(y_i|\theta,\eta;\beta,\sigma_e^2)}{\partial\beta}\, g(\theta,\eta;\, \Sigma_{\theta,\eta})\, d\theta_i\, d\eta_i$$

$$= \sum_{i=1}^{N} \int_\theta \int_\eta \frac{f(\boldsymbol{y_i}|\theta,\eta;\beta,\sigma_e^2)\cdot g(\theta,\eta;\Sigma_{\theta,\eta})}{h_i} \frac{\partial \log f(\boldsymbol{y_i}|\theta,\eta;\beta,\sigma_e^2)}{\partial\beta}\, d\theta_i\, d\eta_i$$

$$= \sum_{i=1}^{N} \int_\theta \int_\eta p_i\, \boldsymbol{X}_i'(\sigma_e^2 \boldsymbol{I}_{\boldsymbol{n_i}})^{-1}(\boldsymbol{y_{it}} - \boldsymbol{X_{it}^T\beta} - \gamma\theta_i - \sigma_u\eta_i)\, d\theta_i\, d\eta_i$$

The equation can be simplified to and set = 0:

$$\frac{\partial logL}{\partial\beta} = \sigma_e^2 \sum_{i=1}^{N} \boldsymbol{X}_i^{'}(\boldsymbol{y_{it}} - \boldsymbol{X_{it}^T\beta} - \gamma\tilde{\theta}_i - \sigma_u\tilde{\eta}_i) = 0.$$

Setting the derivative of $\frac{\partial logL}{\partial\beta}$ equal to 0 results in:

$$\sum_{i=1}^{N} \boldsymbol{X}_i'\, \boldsymbol{X_{it}^T\beta} = \sum_{i=1}^{N} X_i'(\boldsymbol{y_{it}} - \gamma\tilde{\theta}_i - \sigma_u\tilde{\eta}_i).$$

The marginal maximum likelihood solution for the fixed effects covariate vector β is:

$$\hat{\beta} = \left[\sum_{i=1}^{N} \boldsymbol{X}_i'\boldsymbol{X_{it}^T}\right]^{-1} \left[\sum_{i=1}^{N} \boldsymbol{X}_i'(\boldsymbol{y_{it}} - \gamma\tilde{\theta}_i - \sigma_u\tilde{\eta}_i)\right].$$

The fixed-effects covariate matrix β in the LTSPMM the maximum likelihood solution of the random-intercept mixed-effects model where the random effect terms are now modeled by:

$$\tilde{\mu}_{0i} = \gamma\theta_i + \sigma_u\eta_i$$

The empirical Bayes estimate for the mean of the latent trait θ_i is defined as

$$\tilde{\theta}_i = E[\theta_i|\boldsymbol{y_{it}}] = h_i^{-1} \int_\theta \theta_i f_1(\boldsymbol{y_{it}})\, d\theta_i, \qquad (2.47)$$

and the empirical Bayes estimate for the mean of the random intercept η_i is defined as

$$\tilde{\eta}_i = E[\eta_i|\boldsymbol{y_{it}}] = \int_\eta \eta_i f_1(\boldsymbol{y_{it}})\, d\eta_i. \qquad (2.48)$$

The maximum likelihood estimator for the coefficient terms γ of the latent trait θ_i is derived as:

$$\frac{\partial logL}{\partial\gamma} = \sum_{i=1}^{N} \frac{\partial \log h_i}{\partial\gamma}$$

$$= \sum_{i=1}^{N} \frac{1}{h_i} \frac{\partial\left[\int_\theta \int_\eta f(\boldsymbol{y_{it}}|\theta,\eta;\beta,\sigma_e^2)\, g(\theta,\eta;\Sigma_{\theta,\eta})\, d\theta_i\, d\eta_i\right]}{\partial\gamma}$$

$$= \sum_{i=1}^{N} \frac{1}{h_i}\int_\theta \int_\eta \frac{\partial f(\boldsymbol{y_{it}}|\theta,\eta;\beta,\sigma_e^2)}{\partial\gamma}\, g(\theta,\eta;\Sigma_{\theta,\eta})\, d\theta_i\, d\eta_i,$$

and using Equation 2.47 and Equation 2.48 results in:

$$= \sum_{i=1}^{N} \int_\theta\int_\eta \frac{f(\boldsymbol{y_{it}}|\theta,\eta;\beta,\sigma_e^2)\cdot g(\theta,\eta;\Sigma_{\theta,\eta})}{h_i}\frac{\partial \log f(\boldsymbol{y_{it}}|\theta,\eta;\beta,\sigma_e^2)}{\partial\gamma}\, g\ d\theta_i\, d\eta_i$$

$$= \sum_{i=1}^{N} \int_\theta\int_\eta p_i\,(\sigma_e^2\boldsymbol{I_{n_i}})^{-1}(\boldsymbol{y_{it}} - \boldsymbol{X_{it}^T\beta} - \gamma\theta_i - \sigma_u\eta_i)\theta_i'\, d\theta_i\, d\eta_i,$$

and setting the last equation equal to zero results in the marginal maximum likelihood for $\tilde\gamma$,

$$\frac{\partial logL}{\partial\gamma} = \sigma_e^2 \sum_{i=1}^{N} (\boldsymbol{y_{it}} - \boldsymbol{X_{it}^T\beta} - \gamma\tilde\theta_i - \sigma_u\tilde\eta_i)\,\tilde\theta_i' = 0$$

$$\sum_{i=1}^{N} \gamma\tilde\theta_i\tilde\theta_i' = \sum_{i=1}^{N}(\boldsymbol{y_{it}} - \gamma\tilde\theta_i - \sigma_u\tilde\eta_i),$$

which yields:

$$\tilde\gamma = \left[\sum_{i=1}^{N} \tilde\theta_i\tilde\theta_i'\right]^{-1} \left[\sum_{i=1}^{N}(\boldsymbol{y_{it}} - \boldsymbol{X_{it}^T\hat\beta} - \tilde\sigma_u\tilde\eta_i)\right]. \tag{2.49}$$

The maximum likelihood estimator for the coefficient term σu of the standardized random effect η_I is derived as:

$$\frac{\partial logL}{\partial\sigma_u} = \sum_{i=1}^{N} \frac{\partial \log h_i}{\partial\sigma_u}$$

$$= \sum_{i=1}^{N} \frac{1}{h_i} \frac{\partial\left[\int_\theta \int_\eta f(\boldsymbol{y_{it}}|\theta,\eta;\beta,\sigma_e^2)\, g(\theta,\eta;\Sigma_{\theta,\eta})\, d\theta_i\, d\eta_i\right]}{\partial\sigma_u}$$

$$= \sum_{i=1}^{N} \frac{1}{h_i}\int_\theta \int_\eta \frac{\partial f(\boldsymbol{y_{it}}|\theta,\eta;\beta,\sigma_e^2)}{\partial\sigma_u}\, g(\theta,\eta;\Sigma_{\theta,\eta})\, d\theta_i\, d\eta_i$$

$$= \sum_{i=1}^{N} \int_\theta\int_\eta \frac{f(\boldsymbol{y_{it}}|\theta,\eta;\beta,\sigma_e^2)\cdot g(\theta,\eta;\Sigma_{\theta,\eta})}{h_i}\frac{\partial \log f(\boldsymbol{y_{it}}|\theta,\eta;\beta,\sigma_e^2)}{\partial\sigma_u}\, g\ d\theta_i\, d\eta_i$$

$$= \sum_{i=1}^{N} \int_\theta\int_\eta p_i\,(\sigma_e^2\boldsymbol{I_{n_i}})^{-1}(\boldsymbol{y_{it}} - \boldsymbol{X_{it}^T\beta} - \gamma\theta_i - \sigma_u\eta_i)\sigma_u'\, d\theta_i\, d\eta_i,$$

and using Equations 2.47 and Equations 2.48 results in:

$$\frac{\partial logL}{\partial \sigma_u} = \sigma_e^2 \sum_{i=1}^{N} (\boldsymbol{y_{it}} - \boldsymbol{X_{it}^T\beta} - \gamma\tilde{\theta}_i - \sigma_u \tilde{\eta}_i)\ \sigma_u'.$$

The likelihood for σ_u is therefore derived after setting the last equation equal to zero and solving:

$$\sum_{i=1}^{N} \sigma_u\, \eta_i \eta_i' = \sum_{i=1}^{N} (\boldsymbol{y_{it}} - \boldsymbol{X_{it}^T\beta} - \gamma\tilde{\theta}_i),$$

which is solved as:

$$\tilde{\sigma}_u = \left[\sum_{i=1}^{N} \tilde{\eta}_i \tilde{\eta}_i'\right]^{-1} \left[\sum_{i=1}^{N} (\boldsymbol{y_{it}} - \boldsymbol{X_{it}^T\beta} - \gamma\tilde{\theta}_i)\right]. \tag{2.50}$$

The LTSPMM has the following error term e_i defined as $\hat{e}_i = \boldsymbol{y_{it}} - \boldsymbol{X_{it}^T\beta} - \gamma\tilde{\theta}_i - \sigma_u \tilde{\eta}_i$. The maximum likelihood estimator of the error variance σ_e^2 is:

$$\frac{\partial logL}{\partial \sigma_e^2} = \sum_{i=1}^{N} \frac{1}{h_i} \int_\theta \int_\eta \frac{\partial f(\boldsymbol{y_{it}}|\theta,\eta;\, \beta,\sigma_e^2)}{\partial \sigma_e^2}\, g(\theta,\eta;\, \Sigma_{\theta,\eta})\, d\theta_i\, d\eta_i$$

$$= \sum_{i=1}^{N} \int_\theta \int_\eta \frac{f(\boldsymbol{y_{it}}|\theta,\eta;\, \beta,\sigma_e^2) \cdot g(\theta,\eta;\, \Sigma_{\theta,\eta})}{h_i} \frac{\partial \log f(\boldsymbol{y_{it}}|\theta,\eta;\, \beta,\sigma_e^2)}{\partial \sigma_e^2}\, d\theta_i\, d\eta_i$$

$$= \int_\theta \int_\eta p_i \left[-\frac{n_i}{2}\sigma_e^2 + \frac{1}{2}\sigma_e^{-4}(\boldsymbol{y_{it}} - \boldsymbol{X_{it}^T\beta} - \gamma\theta_i - \sigma_u\eta_i)' (\boldsymbol{y_{it}} - \boldsymbol{X_{it}^T\beta} - \gamma\theta_i -\right.$$

$$\left.\sigma_u\eta_i)\right]\, d\theta_i\, d\eta_i$$

$$= \frac{1}{2}\sigma_e^{-4} \sum_{i=1}^{N} \left(-n_i\, \sigma_e^2 + e'e + tr\left|\gamma\sigma_u \Sigma_{\theta,\eta|\boldsymbol{y_{it}}} \gamma\sigma_u'\right|\right) = 0.$$

Shared Parameter with Logistic Regression

X. Lin et al. (2018) introduced a shared parameter location scale model to research the informative and intermittent missingness with respect to both the mean and within subject variation of the primary outcomes. The shared parameter model assumes there is a set of latent variables U_i shared between the primary outcome and missing process, which are conditionally independent given U_i. The model extends the work of Cursio et al. (2019) by allowing a random subject scale effect to influence missingness while implementing a Bayesian model estimation framework instead of maximum likelihood methods.

The location and scale random effect model (Hedeker et al., 2008) is implemented to characterize the heterogeneity of the primary outcomes. The random intercept location parameter for the participants is as follows

$$y_{it} = \beta_0 + \boldsymbol{X}_{\boldsymbol{it}}^{\boldsymbol{T}}\boldsymbol{\beta} + u_{0i} + e_{it}. \tag{2.51}$$

Let y_{it} be the outcome for participant i at occasion t, where $i = 1, \dots, n$, and $t = 1, \dots, k_i$ allowing the participants to have different number of measurements by subscript k with i. The fixed effect covariate vectors are $\boldsymbol{X}_{\boldsymbol{it}}$ and can include subject and occasion level covariates. The random location effect is u_{0i} indicating the effect of participant i on his/her mean of the repeated measurements and is normally distributed as $u_{0i} \sim N(0, \sigma_u^2)$. The random error term e_{it} reflects the uncertainty in measuring participant i's outcome at occasion t relative to the participant average and is normally distributed as $e_{it} \sim N(0, \sigma_e^2)$. The random scale effects extension for the within subject variance model is

$$\log\left(\sigma_{e_{it}}^2\right) = \alpha_0 + \boldsymbol{Q}_{\boldsymbol{it}}\boldsymbol{\alpha} + u_{2i}, \tag{2.52}$$

where the log function ensures that the estimated error variance is strictly positive. The term Q_{it} is the within subject variance design matrix and usually contain a subset of the variables in $\boldsymbol{X}_{\boldsymbol{it}}$. The $\boldsymbol{\alpha}$ is the fixed effect coefficient vector and indicate the effect of the within subject coefficient on the log-variance of the outcome. The random subject scale intercept is u_{2i} indicating the effect of participant i on his/her within subject variability of the repeated measurements.

The model for the missing process uses a random intercept logistic regression model for the binary missing prompt indicators. The responsiveness R_{it} represents the responding indicator for participant i at occasion t, where R_{it} is 1 if the participant responds to the prompt and 0 if participant missed the prompt. The random intercept logistic regression model is given by

$$log\left(\frac{\Pr(R_{it}=1)}{1-\Pr(R_{it}=1)}\right) = \tau_0 + \sum_{K=n}^{K=2} \tau_k * T_{it}^{k} + \lambda_i, \tag{2.53}$$

where $k = 2, \dots, n$ is the time-bin index and T_{it}^{k} is the indicator of the kth time-bin for the prompt individual i received at occasion t. For example, if a study prompted participants three times a day and starts on Monday morning, the first time-bin T_{i1}^{1} is treated as the reference time-bin. The fixed intercept is τ_0, indicating the log odds of missing a response for an individual with $\lambda_i = 0$. The term λ_i is participant i's random intercept, indicating the influence of subject i on his/her log odds of missing prompts and follows a normal distribution of $\lambda_i \sim N(0, \sigma_\lambda^2)$. Conditional on λ_i, the responding indicators R_{it} are assumed to be i.i.d following a Bernoulli distribution with missing probability:

$$P_{it} = \frac{\exp(\tau_0 + \sum_{K=n}^{K=2} \tau_k * T_{it}^{k} + \lambda_i)}{1+\exp(\tau_0 + \sum_{K=n}^{K=2} \tau_k * T_{it}^{k} + \lambda_i)}. \tag{2.54}$$

Here P_{it} is modelled by both observed and latent information, with time-bins being explicitly measured and the random effect λ_i accounting for all unobserved information at the participant level.

The joint model combines the outcome model, the dispersion model, and the missing process. The random subject effects for the random intercept location u_{0i} and the random scale effect u_{2i} in the outcome and missing model leads to the parameter sharing of λ_i displayed in the equations below:

$$u_{0i} = \gamma\lambda_i + \eta_{0i} \tag{2.55}$$

$$u_{2i} = \delta\lambda_i + \eta_{2i} \tag{2.56}$$

where $\{u_{0i}, u_{2i}\}$ and λ_i are traits specific to individual i with a set of linear models used to link the random effects. Adding the shared parameter addition to the random location effect u_{0i} to

Equation 2.51 and random scale effect u_{2i} in Equation 2.52 adds the additional missing parameters to the model:

$$y_{it} = \beta_0 + \boldsymbol{X_{it}^T\beta} + (\gamma \cdot \lambda_i + \eta_{0i}) + e_{it} \tag{2.57}$$

$$\log(\sigma_{e_{it}}^2) = \alpha_0 + \boldsymbol{Q_{it}\alpha} + (\delta \cdot \lambda_i + \eta_{2i}). \tag{2.58}$$

In Equation 2.57, individual i's location random intercept u_{0i} is modeled using trait λ_i and an error term η_{0i} for each participant that is normally distributed $\eta_{0i} \sim N(0, \sigma_\eta^2)$. The coefficient γ indicates the effect of missingness on the participant's mean outcome. In Equation 2.58, individual i's random scale effect u_{2i} is modeled by missing trait λ_i and an error term η_{2i} for each participant that is normally distributed $\eta_{2i} \sim N(0, \sigma_\eta^2)$. The coefficient δ indicates the effect of missingness on the within-subject variability of the outcome. Informative missing is accounted for by linking the missing process through the random effects and the outcome. X. Lin et al. (2018) termed λ_i as the shared random subject effect between the outcome and missingness, and η_{0i} and η_{2i} as the residual random subject location and scale effects.

X. Lin et al. (2018) conducted a series of simulation studies where observations were set to intermittent missing under two scenarios: (1) missing does not depend on the potential outcomes (MCAR or MAR), and (2) missing depends on potential outcomes (MNAR). They compared the shared random subject location, scales effects between the outcome and missing process against a naïve LMM that implements list-wise deletion and utilized only the observed outcome. Under the MAR missing data situation, the two models performed similarly. Under an MNAR missing data situation, shared parameter model had smaller bias and better coverage rate for the within subject intercept and gender coefficients than list-wise deletion. List-wise deletion ignore the association between the outcome and missing process that can lead to invalid inferences. X. Lin et al. (2018) concluded that the shared parameter model achieves good estimation precision, correct interval length and asymptotic coverage rate for the computed parameter estimates yet provides insightful information about the missing mechanisms when the missing data is MNAR. The findings displayed evidence that the shared parameters for the random subject location and scale effects have an association between the missing process and the outcome.

Model Estimation

Bayesian estimation for the shared parameter model denote ϕ = (β, α, τ, γ, δ,) as the model parameter vector, $\lambda = \{\lambda_i\}_{i=1}^{n}$ as the random subject effects for the missing process, $\eta = \{\eta_{0i}, \eta_{2i}\}_{i=1}^{n}$ as the random subject effect vector in the outcome model and $D = \{Y_i, M_i\}_{i=1}^{n}$ as the data. Parameters ϕ, λ, and η are regarded as random and follow some prior distribution before we get to observe the data D, which are denoted as $\pi(\phi)$, $\pi(\lambda)$, and $\pi(\eta)$ respectively. Univariate standard normal and bivariate standard normal are choices for $\pi(\lambda)$ and $\pi(\eta)$. For $\pi(\phi)$, one can specify a separate prior for each component in ϕ provided that a full conditional posterior is obtained for each of them. Given independent priors, one can derive the conditional posterior as

$$P(\phi|\lambda_i, \eta_i, D_i) \propto P(D_i|\phi, \lambda_i, \eta_i)\pi(\phi), \tag{2.59}$$

$$P(\lambda_i|\phi, \eta_i, D_i) \propto P(D_i|\phi, \lambda_i, \eta_i)\pi(\lambda_i), \tag{2.60}$$

$$P(\eta_i|\phi, \lambda_i, D_i) \propto P(D_i|\phi, \lambda_i, \eta_i)\pi(\eta_i). \tag{2.61}$$

$P(D_i|\phi, \lambda_i, \eta_i)$ is the conditional joint likelihood and π is the corresponding prior. Once the full conditional posteriors are obtained for ϕ, *λ, and* η, the joint posterior can be approximated by sampling each variable from its full conditional posterior iteratively using Gibbs sampling (Casella & George, 1992).

Chapter II Summary

In summary, missing data in longitudinal and ILD studies are common and correctly implementing missing data methods can lead to unbiased and efficient parameter estimate improving the nature of variable relationships and correct conclusions. Longitudinal studies have a history of research with some proven methods for the many diverse missing data situations. ILD or EMA studies have grown in the last 20 years and have less research on methods for missing data. The challenge researchers' face is how to manage the intermittent missing data on high volume of assessments. There is much needed research in the literature on missing data in ILD studies. Motivated by the need of more research on this topic, I propose to study the shared parameter missing models of Cursio et al. (2019) and X. Lin et al. (2018) more extensively to fill a gap of missing data methods for ILD studies.

CHAPTER III

METHODOLOGY

The following list is a reminder of the research questions that will be answered by comparing the performance of the two shared parameter missing data models.

Q1 Which model, ILD missing data models LTSPMM (Cursio et al., 2019), SPLR (X. Lin et al., 2018), or the full mixed-effect location random effects model using list-wise deletion, perform better under different combinations of number of prompts (25, 40) and intermittent missingness scenarios (20%, 30%) in terms of raw bias percentage?

Q2 Which model, ILD missing data model LTSPMM (Cursio et al., 2019), SPLR (X. Lin et al., 2018), or the full mixed-effect location random effects model using list-wise deletion perform better under different combinations of number of prompts (25, 40) and missingness scenarios (20%, 30%) in terms of empirical standard errors?

Q3 Which ILD missing data model LTSPMM (Cursio et al., 2019) or the SPLR model (X. Lin et al., 2018), performs more computationally efficient in terms of computational run time?

Introduction

ILD and EMA studies repetitively collect assessments on participants in their real-world environments focusing on their current states providing information about how their behaviors and experiences vary over time and across situations. Collecting information in the moment aims to avoid bias and errors associated with recall. EMA studies randomly send participants several signaled prompts on a portable device or APP on their cell phone over the course of hours, days and weeks allowing enough time for a representative sample to answer their research questions. The distractions and complexities of daily life leads to some participants missing responses. For example, participants may miss assessments because they are in class or have an unexpected meeting and are unable to respond to the prompting device. Missing data through non-compliance can have a significant effect on statistical power, but also conclusions that can be drawn through statistical inference (Graham, 2009). The participants typically miss entire assessments intermittently throughout the course of the study. The intermittent missing data is complex making longitudinal missing data models limited in certain cases. Silvia et al. (2013) stress that examining nonresponse is critical for EMA research. There is a need for more research on missing data models intended to investigate nonresponse of assessments in EMA studies.

The joint shared parameter missing data models by (Cursio et al., 2019) and (X. Lin et al., 2018) exhibited new ways to analyze and estimate EMA data containing problematic intermittent missing data. The two shared parameter models were adapted for ILD where intermittent and informative missing happen often due to missed prompts, making it possible to perform valid statistical inference. The missing data methods have similarities as they both treat participants missing prompts as a latent variable by introducing a shared parameter to the location model but do so in different ways. The LTSPMM implements an IRT logistic model that estimates a latent variable describing participants' ability to respond to the prompting device. The shared parameter location scale model utilizes a logistic regression model for the binary missing prompt indicators that estimates the log odds of response. The purpose of this study was to compare these two joint models used to handle missing data in ecological momentary assessment (EMA) studies and to evaluate their performance under different assessment and missing data scenarios. Motivated by the concepts of the X. Lin et al. (2018) SPLR and the LTSPMM (Cursio et al., 2019), the goal of this research is to compare the two models from a statistical perspective under different EMA prompt designs with varying levels of participant intermittent missingness conditions. I will illustrate a series of simulation studies as well as real data application using these missing models assessing missing data in terms of raw bias, empirical standard errors (SE) and computation run time. A holistic view that combines run time, raw bias percent, and empirical standard errors will provide valuable knowledge about the limitations and capabilities of these missing data methods. For example, a method may reduce computation run time but have more raw bias percent with larger empirical standard errors making it more difficult to trust. The next sections will give details on the empirical study, simulation plan, and a breakdown of research questions one through three.

Empirical Study

A real data situation with intermittent missing ILD values will be used with the candidate models by Cursio et al. (2019) LTSPMM, X. Lin et al. (2018) SPLR location only model, and a full mixed-effect location random effects model using list-wise deletion. The data described in this section for the empirical study was provided by Phillips et al. (2015) contributing theoretical information on the deterministic model for the simulation. One of the aims of the empirical study was to examine the association of marijuana craving and its relationship to academic motivation when assessed in the moment with college students that tested positive for using marijuana. The goals explored if heavy marijuana users craving is associated with less time spent academically on tasks like studying and managing academic goals. Phillips et al. (2015) in the moment study found that craving was negatively associated with academic effort and motivation.

The participants (*n* = 110) completed the EMA prompts via the Reallife Exp application (lifedatacorp.com) over 14 consecutive days. Participants were prompted by a notification thru the APP on their smart phone three times a day with one prompt randomly falling within each of the following strata: first notification transpired between 8:00 a.m.-12:00 p.m. (morning), second notification occurred between 12:30-4:30 p.m. (afternoon), and third notification took place between 5:00-9:00 p.m. (evening). Each participant received 42 prompts during their time in the study. The EMA questions focused on the participants' current activity, academic motivation, craving for marijuana, anxiety, mood, marijuana use and frequency of use since last prompt, social setting when using, exercise, number of alcoholic drinks since last prompt, and learning behaviors like time spent studying.

After recruitment, the participants completed a series of baseline self-report measures. The baseline measure that will be included for this study is the psychometric questionnaire the Rutgers Marijuana Problem Index (RMPI, White et al., 2005). The 23-item version assesses negative consequences associated with marijuana use within the last year. Items are rated from 0 to 3 ("none" to "more than 5 times") based on the frequency of each consequence. In this study, the dependent variable will be craving level that was collected when participants responded to the prompts during their 14 days in the study. Craving is described as a strong intense urge or desire to use marijuana and is captured in the moment for each of the forty-two EMA assessments. Participants were asked to rate their current marijuana craving at this exact moment on a scale of 0 to 10. This 11-point scale ranged from 0 "no cravings" to 10 "extremely intense cravings." Participants were asked to rate their academic motivation in the moment for each assessment. Academic motivation is defined as paying attention in college courses, completing reading and homework assignments, and studying. Participants were asked how motivated they currently feel to focus on schoolwork? This scale ranged from 0 to 10, with 0 being "not at all" and 10 being "extremely motivated."

The participants responded to 3,697 of the possible 4,620 prompts resulting in an 80% response rate. Only two of the participants responded to all 42 prompts. Five participants responded to less than half of the prompts with the lowest responder in the study answering only nine times. The overall participant average responses to the prompting device were 33.6 prompts with a standard deviation of 6.1. Of the three daily notifications, the morning notification from 8:00 a.m.-12:00 p.m. had the most missing prompts occur with 397, accounting for 43% of the total missing responses. The missing responses for the afternoon notification that occurred between 12:30-4:30 p.m. were 255 and 27.6% of the total. The evening notification from 5:00-9:00 p.m. had similar total missing responses to the afternoon notification with 271 that was 29.4% of the total. The highest total of missing for any of the possible notification times and days in the study occurred on Sunday morning for the 8:00 a.m.-12:00 p.m. notification with 91 missing responses which results in 9.8% of the total missing responses. The participants in this study responded less often on the weekends with 323 total missing responses on Saturday and Sunday resulting in 35% of the total missing.

Simulation Plan

A simulation will be conducted to answer all the research questions by comparing the intermittent missingness on an intensively measure longitudinal outcome using the LTSPMM (Cursio et al., 2019), SPLR location only model (X. Lin et al., 2018), and a full mixed-effect location random effects model using list-wise deletion missing models. List-wise deletion will be used as a comparison model as it is a default missing data method on many statistical packages including SAS (SAS Institute Inc, 2013) and R (R Core Team, 2021). The purpose of the simulation is to compare a true model with no missing data to the selected missing methods under different numbers of assessment (25, 40) and introducing varying levels of percentages of intermittent missingness (20%, 30%) in terms of raw bias percentage, empirical standard errors (SE) and computational run time. In the next section, a description of the simulation design will be outlined, followed by the process.

Simulation Statistical Software

The data will be simulated using the statistical software R (R Core Team, 2021) and the package "*simglm*" (LeBeau, 2023). The "*simglm*" function features simulations of multi-level longitudinal data allowing for users to specify the distribution of the random components. The package can simulate ordinal variables like the academic motivation and marijuana craving 11-point scales allowing for empirical data information to include means and standard deviations. The function includes the option to add time-dependent covariates like academic motivation in the proposed model that will allow for within and between participant analysis. Last and the most important feature for this package is generating random intermittent missing data by percentages in the form of an indicator variable allowing the ability to check that the missing data was generated properly. The proportion missing will randomly be introduced using an indicator variable with a 1 signifying a prompt that is missing and a 0 signifying an answered prompt. The missing data variable provides a convenient way to analyze the full data set with data sets that have missing data. The features of "*simglm*" (LeBeau, 2023) provide all the necessary elements for a successful simulation on the two share parameter models.

Sample Size

Sample sizes in ILD research are dependent on the topic and the available populations to study. Jones et al. (2019) displayed the variance in sample size for substance use ILD studies with a mean of 154.21, standard deviation of 214.8 with participant sizes ranging between 10 and 1021. The meta-analysis included data collections that were conducted using PDAs, smartphones, or internet and sample sizes per type was not included. Data collection for ILD research using smartphones is the likely feasible option and is expected to be the preference for researchers as it was estimated in 2019 that 3.5 billion people own a smart phone, which was one-third of the population including 70% of people in the Western world and United States of America (GSMA intelligence, 2019). Smartphones allow for flexible designs with easy data collection and researchers will not have to supply participants with devices. In a more recent meta-analysis on smartphone use for 53 ILD studies the median sample size reported was an n=97 participants (de Vries et al., 2021). With the high variance of sample sizes amongst ILD studies using a sample size close to the median for smartphones research represents a logical target for this simulation study. The empirical study had an n=110 participants so a sample size of 100 participants will be implemented for all missing and assessment designs.

Missing Data Setup and Percentages

The missing data setup for the simulation will be randomly intermittent missing which is the typical type for ILD studies and were discussed in the chapter two sections intensive longitudinal data and patterns of missing data. The ILD missing data situations are unsuitable to implement Rubin (1976) and colleagues (Little & Rubin, 2002) missing data mechanisms as the complicated nature of missingness is hard to specify. Cursio et al. (2019) indicated that the assumptions about the nature of the missing data are typically unknown in EMA studies, and, in many cases, the missing data is complex and highly irregular. X. Lin et al. (2018) tested their shared parameter missing model against simulated MAR and MNAR data but warned that one cannot know whether data are missing at random or not in practice when the true underlying missing mechanism is unknown. The missing data situation researchers are facing is at the discretion of the individuals in the study responding to prompts sent to them. Silvia et al. (2013)

found that the missing data was at the within-person level. Many ILD study designs are set up to reduce missing data amongst the participants in the study by avoiding specific schedules participants face in order to make the project as successful as possible. For example, Phillips et al. (2015) set up individualized prompting designs so that the college students did not receive prompts while they were in class. Missing data patterns could be different depending on prompting designs, number of prompts, or the type of individuals in the study. Wrzus and Neubauer (2023) found that the total number of assessments, the number of assessment days, or the number of assessments per day did not predict participants' compliance with the assessment schedule. However, missing data will occur at the individual level in all ILD studies. Therefore, the prompted responses will be considered MCAR and simulated randomly intermittent missing amongst the participants based on a condition percentage.

Multiple meta-analysis for EMA literature indicates that social researchers report average compliance rates to the prompting device between seventy and eighty percent with the number of assessments per day ranging from two to nine on various human behaviors. Wen et al. (2017) reviewed studies using mobile devices to collect EMA data among youth (age ≤18 years old) and found an overall compliance rate of 78.3%. Liao et al. (2016) reviewed studies addressing nutrition and physical activity in youth reporting an average compliance rate of 71%. Wrzus and Neubauer (2023) examined a wide range of topics like health behaviors, mental health, emotions, social relationships and others for adults and youth and found overall compliance rates of 79.19%. Finally, Jones et al. (2019) reviewed studies related to substance use and reported a pooled compliance rate of 75.06%. Implementing missing data percentages like what researchers may encounter will make the simulation more applicable. This simulation will set the missing percentages to 20% and 30% using the missing data feature in the "*simglm*" (LeBeau, 2023) function making them comparable to real ILD data situations.

Assessment Design

The literature displays that researchers have a variety of assessment designs for ILD studies with the goal of collecting events and experiences in an individual's life. In ILD, one assesses moments or periods of time, raising the issue of how to ensure that the moments or periods assessed are representative of the subject's experience (Shiffman et al., 2008). The scheme for frequency and timing including respect for subject burden can implement countless schemes depending on the goals of the study. The decision about how many samples or recordings per day are needed should also be guided by the nature of the phenomenon to be recorded (Stone & Shiffman, 2002). For example, a fourteen-day assessment period was a time frame that has been demonstrated to be adequate to assess substance use behaviors in past EMA studies (Buckner et al., 2012; Phillips et al., 2015; Shrier et al., 2012). While there are many different assessment designs researchers are implementing, compliance across all the different designs is stable. The total number of assessments, the number of assessment days, or the number of assessments per day did not predict participants' compliance with the assessment schedule (Jones et al., 2019; Wrzus & Neubauer, 2023). Choosing a perfect assessment design for all researchers is near impossible and since literature reveals response rates are similar across all designs, implementing a prompting design like the empirical study will be the aim of this simulation. Recall that Phillips et al. (2015) assessments were designed to prompt the members three times daily stratified across morning, afternoon, and evening over the course of two full weeks or fourteen continuous days for a total of forty-two participant assessments. Other ILD designs might require less time needed to study specific behaviors. Wrzus and Neubauer (2023) offered a wide range of researched topics using ILD schemes and reported that the impression arises that most studies ask participants to report around 5 times a day for a week or less. An additional design related to approximately half of the empirical study comparable to one-week designs will provide more information for ILD researchers on missing data situations with less scheduled assessments. The assessment designs conditions chosen for this simulation are 25 and 40.

Empirical Data for Simulation

The research conducted by Phillips et al. (2015) demonstrated empirical evidence of a statistically significant relationship between marijuana craving and academic motivation amongst college students that tested positive for marijuana. The collection of real-world data provides valuable in the moment behaviors of the college students while navigating the rigors of schoolwork and life at a university. The empirical data captures the relationship from the observational research and provide the foundation for the simulation. The "*simglm*" (LeBeau, 2023) package allows for the specification of random components therefore implementing a full mixed-effect location random effects model from the empirical data will provide information needed for the simulation. The outcome variable of marijuana craving is assumed to be normally distributed $Y_{it} \sim N(2.86\ 2.87)$ and equation 3.1 provides an example of this model:

$$Craving_{it} = int + Time_{it} + RMPI_i + Motiv_w + Motiv_B + u_{0i} + e_{it}. \tag{3.1}$$

Results of the estimated mixed effect model of the empirical data provide information about the random intercept variance u_{0i} will be set to 12.579 and the random residual effect variance e_{it} set to 5.593. Information about the variables in the simulation will be derived from their means and standard deviations from the empirical data. The intercept int will be set to the value 2.41. The time variable $Time_{it}$ displays an average increase of cravings amongst the participants over time in the study and will have a value of .007921 that will be included as a level 1 variable. Academic motivation will be added as an ordinal variable on an11-point scale ranged from 0 “not at all” and 10 being “extremely motivated.” This is the time dependent variable which means $Motiv_w$ will be included as level 1 variable and $Motiv_B$ will be included as a level 2 variable. A negative association will be established by the correlation of the motivation and craving variables which is -.07. The distribution will be set to the probability of response from the members in the empirical study for all the levels of the motivation ordinal variable from 0 to 10. For example, 15% of the members responded 0 motivation thus the probability of response will be set to .15. The Rutgers Marijuana Problem Index $RMPI_i$ will be added as an ordinal variable ranging from 0 to 69 possible scores. This variable will be a level 2 variable and distributed random normal with a mean of 13.95 and standard deviation of 9.32 for the participants in the study.

Hypothesized Model

A hypothesized model with no missing data will be established for each iteration of the simulation using intensively measured longitudinal outcome craving (Y_{it}) for (100) individuals at two different levels of assessment prompts 25 and 40. Covariates included in the model will be the RMPI (subject level 2), time-dependent covariate will be the motivation variable (subject level 2 and occasion level 1), and time stamps (occasion level 1). After the hypothesized model is generated using variable information from the empirical study, observations will set the prompt response to intermittent missing via “*simglm”* (LeBeau, 2023) for each of the assessment levels at two different missingness percent levels 20% and 30%. A hypothesized model with no missing data will be estimated for all four different simulation conditions and will be compared after applying the missing data models LTSPMM (Cursio et al., 2019), SPLR location only model (X. Lin et al., 2018), and a full mixed-effect location random effects model using list-wise deletion after introducing missing data conditions. The bias for the proposed models with missing data helps examine the performance for each parameter as the average point deviations from the true value. The estimated hypothesized model with no missing data T for each of the parameters P will be determined for each condition c for the parameter estimates of the full data set with no missing data as follows:

$$Hypothesized\ Model = P_c^T, \tag{3.2}$$

where parameters $P = (\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_B, \hat{\beta}_w, \hat{u}_{0i})$ are established for all the conditions $c = 1, \ldots ,4$. The three intensively measured longitudinal and corresponding missing model specifications will be displayed next followed by the simulation process.

Proposed Models

The proposed models for the simulation will compare the estimates with missing data analyzing the following missing data methods of a full mixed-effect location random effects model using list-wise deletion (Equation 3.3), X. Lin et al. (2018) SPLR location only model (Equation 3.5), and Cursio et al. (2019) LTSPMM (Equation 3.7) to the hypothesized model. For all the candidate models Y_{it} denotes the marijuana craving outcome that is assumed to be normally distributed $Y_{it} \sim N(0, \sigma_Y^2)$. In all the models, β_1 represents the average effect of time on the marijuana craving level. The estimated regression coefficient β_2 represents the between-subject covariate baseline scale Rutgers Marijuana Problem Index (RMPI). The most important covariate for this study is the time-dependent academic motivation with β_B denoting the regression coefficient for the academic motivation between-subjects as a level 2 covariate and β_w is the regression coefficient within-subject as a level 1 covariate. Notice that the full mixed-effect location random effects model using list-wise deletion (Equation 3.3) does not have a missing model. List-wise deletion is the default missing data method for R (R Core Team, 2021) and SAS (SAS Institute Inc, 2013) and is amongst the most commonly used missing data methods in literature (Lang & Little, 2018; Peugh & Enders, 2004). This will act as the worst-case method for handling missing data in this simulation. The shared parameter missing models found in Equations 3.4 and 3.6, the coefficient γ indicates the effect of missingness on the participant's mean marijuana craving level and η_{0i} is the remaining random location effect term for participant i on his/her mean of the repeated measurements.

The full mixed-effect location random effects model using list-wise deletion:

$$Y_{it} = \beta_0 + \beta_1 X_{it} + \beta_2 X_i + \beta_w (X_{it} - \bar{X}_{i.}) + \beta_B \bar{X}_{i.} + u_{0i} + e_{it}, \tag{3.3}$$

X. Lin et al. (2018) SPLR location only missing and random effect models:

$$log\left(\frac{\Pr(R_{it}=1)}{1-\Pr(R_{it}=1)}\right) = \tau_0 + \sum_{K=n}^{K=2} \tau_k * T_{it}^{k} + \lambda_i, \tag{3.4}$$

$$Y_{it} = \beta_0 + \beta_1 X_{it} + \beta_2 X_i + \beta_w (X_{it} - \bar{X}_{i.}) + \beta_B \bar{X}_{i.} + \gamma \lambda_i + \eta_{0i} + e_{it}, \tag{3.5}$$

Cursio et al. (2019) LTSPMM one-parameter missing and random effect models:

$$log\left(\frac{\Pr(R_{it}=1|\theta_i)}{1-\Pr(R_{it}=1|\theta_i)}\right) = \frac{1}{1+\exp[-a(\theta_i - b_t)]}, \tag{3.6}$$

$$Y_{it} = \beta_0 + \beta_1 X_{it} + \beta_2 X_i + \beta_w (X_{it} - \bar{X}_{i.}) + \beta_B \bar{X}_{i.} + \gamma \theta_i + \eta_{0i} + e_{it}. \tag{3.7}$$

Simulation Process

This section will outline the simulation process in a step-by-step way displaying the different simulation settings.

1. Generate an intensive longitudinal data using R "*simglm"* (LeBeau, 2023) for 100 participants with 25 and 40 assessments for a total of 2,500/4,000 data points.
2. Set observations to 20% and 30% intermittent missing via "*simglm"* (LeBeau, 2023) for both assessment conditions resulting in four total combinations.

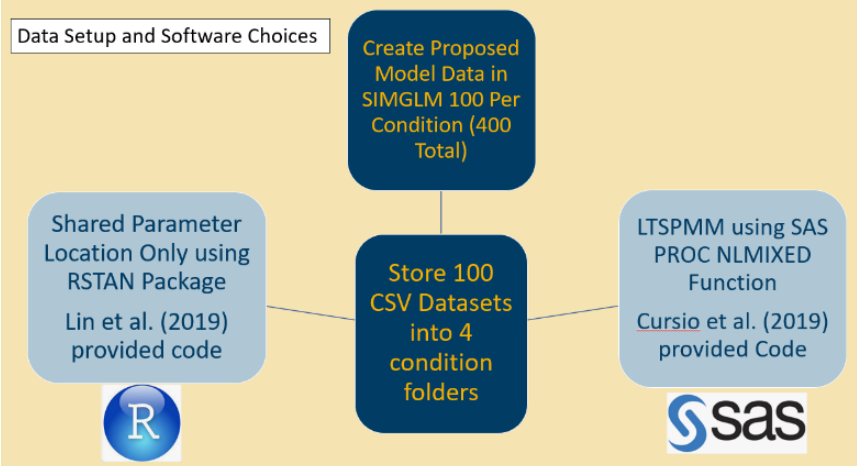
3. Conduct analyses using the candidate methods of the full mixed-effect location random effects model using list-wise deletion model, shared parameter with logistic regression missing model, and LTSPMM saving the run time, βs, location random effects u_{0i}, shared parameter coefficients γ.
4. Repeat 100 times (X. Lin et al., 2018), for each combination of prompts 25/40 and intermittent missing percentages 20% and 30%.
5. Calculate bias using the formula Raw Bias Percent $= \frac{P^T_{cp} - \overline{P}^s_{mcp}}{P^T_{cp}}$.
6. Calculate Standard Errors using the formula $EmpSE = \sqrt{\frac{1}{m-1}\sum_{k=1}^{m}(\hat{P}^s_{kcp} - \bar{P}^s_{mcp})^2}$.
7. Calculate the average and find the median of run time with 95% confidence intervals.

Data Setup and Software Choices

Comparing the two shared parameter missing data models will have a unique setup for the simulations. The reason being both authors of the two shared parameter missing data models derived and provided code from different statistical software packages. Keeping the integrity of their code is important for this simulation. X. Lin et al. (2018) provided code using the rstan package (Stan Development Team, 2022) within the R statistical package (R Core Team, 2021). Cursio et al. (2019) derived code from SAS (SAS Institute Inc, 2013) using the PROC NLMIXED function. Therefore, all the combination of prompts and intermittent missing percentages will be created and stored as individual CSV (comma-separated values) files in four separate master folders. The result will be one hundred data files stored in each of the four folders, which was described in step four above in the simulation process section. To analyze the data simulation code will be setup separately in R and SAS that will load each file and execute the code saving all the estimates described in step 3 of simulation process section to compare the hypothesized model with the candidate models. Figure 2 displays a visual of the unique setup for the simulation.

Figure 2

Diagram of the Data Storage and Simulation Process


Data Setup and Software Choices
Create Proposed Model Data in SIMGLM 100 Per Condition (400 Total)
Shared Parameter Location Only using RSTAN Package
Lin et al. (2019) provided code
Store 100 CSV Datasets into 4 condition folders
LTSPMM using SAS PROC NLMIXED Function
Cursio et al. (2019) provided Code
R
sas

Research Question 1 Analysis Plan

Q1 Which model, ILD missing data models LTSPMM (Cursio et al., 2019), SPLR (X. Lin et al., 2018), or the full mixed-effect location random effects model using list-wise deletion, perform better under different combinations of number of prompts (25, 40) and intermittent missingness scenarios (20%, 30%) in terms of raw bias percentage?

To compare the missing models of the LTSPMM, shared parameter location only, and the full mixed-effect location random effects model using list-wise deletion under different assessment and missing data conditions, performance in terms of raw bias percentage will be evaluated from the estimated unstandardized regression coefficients, $\hat{\beta}$. Raw bias percentage will be computed for each parameter as the average point deviation from the true value of the true model with no missing data. The bias will help quantify how well the missing data methods are estimating the parameters on average with a key property being unbiasedness. The raw bias percent provides information about the accuracy of the coefficient estimates about the true model and the three missing data methods on performance of missing data fit in the simulated data sets. First, the average parameter P estimate from each of the simulations s will be computed as:

$$\frac{1}{100}\sum_{k=1}^{100} P_{kcp}^{s} = \bar{P}_{mcp}^{s}. \tag{3.8}$$

The estimated simulation coefficient P^{s} for simulation $k = 1, \dots ,100$, at condition $c = 1, \dots ,4$, and for parameter $P = (\hat{\beta}, \hat{u}_{0i}, \hat{\gamma})$. Raw bias percentage is the true parameters from the corresponding non-missing dataset minus the parameter estimates averaged across all converged replications as follows:

$$Raw\ Bias\ Percent = \frac{P_{cp}^{T} - \bar{P}_{mcp}^{s}}{P_{cp}^{T}}. \tag{3.9}$$

Raw bias percent will be computed for the parameters $p = (\ \hat{\beta}_0, \hat{\beta}_B, \hat{\beta}_w)$ for all the different assessment and missingness conditions $c = 1, \dots ,4$. The variables of interest for this simulation is the time dependent variable that represents the within subject motivation $\hat{\beta}_w$ at level 1 associated with the slope and between subject motivation $\hat{\beta}_B$ at level 2. The parameters bias will be evaluated for all three missing data models and provide information about their overall performance for all the various assessment and missing conditions.

Research Question 2 Analysis Plan

Q2 Which model, ILD missing data model LTSPMM (Cursio et al., 2019), SPLR (X. Lin et al., 2018), or the full mixed-effect location random effects model using list-wise deletion perform better under different combinations of number of prompts (25, 40) and missingness scenarios (20%, 30%) in terms of empirical standard errors?

To compare the missing models LTSPMM, shared parameter location only, and the full mixed-effect location random effects model using list-wise deletion under different assessment and missing data conditions, evaluating empirical standard errors (SE) from the estimated unstandardized regression coefficients of $\hat{\beta}$. The empirical SE estimates the long-run standard deviation of the parameters for all the simulated data sets. Thus, it is a measurement of precision or efficiency of the estimators for the missing data models. Computing empirical SE is accomplished by taking the regression coefficient obtained for each parameter and condition from each simulation minus the parameter average estimate of that condition as follows:

$$EmpSE = \sqrt{\frac{1}{m-1}\sum_{k=1}^{m}(\hat{P}_{kcp}^{s} - \bar{P}_{mcp}^{s})^{2}} \qquad (3.10)$$

where estimated simulation coefficient P^{s} for simulation $k = 1, \dots ,100$, at condition $c = 1, \dots ,4$, for parameter $P = (\,\hat{\beta})$. A high empirical SE indicates that the missing data method tends to produce highly varied results from sample to sample.

Research Question 3 Analysis Plan

Q3 Which ILD missing data model LTSPMM (Cursio et al., 2019) or the SPLR model (X. Lin et al., 2018), performs more computationally efficient in terms of computational run time?

One challenge to the application of joint models is its computational complexity (Hickey et al., 2016; Yang et al., 2016). This is a concern for practical use in both the LTSPMM and the SPLR location only missing model. X. Lin et al. (2018) proposed a full Bayesian estimation approach that was described as computationally demanding but did not mention computational run time or convergence issues. Cursio et al. (2019) implemented maximum marginal likelihood estimation where all joint models took at least 3 hours with reported convergence issues. Computational run time for the joint missing models will provide information on how practical the missing data methods are under different assessment and missingness conditions. To look at the difference between computational runtimes between the LTSPMM and the shared parameter location scale missing model averages, medians, and confidence intervals. Computing the runtime R from each of the simulations s is

$$Average\ Runtime = \frac{1}{100}\sum_{k=1}^{100} R_{kcl}^{s} = \bar{R}_{cl}^{s}. \qquad (3.11)$$

The simulation run time average is calculated as R^s for simulation $k = 1, \dots, 100$, at condition $c = 1, \dots, 4$ for missing models *l = (LTSPMM, Shared Parameter with Logistic Regression)*. The median of the runtimes of 100 simulations will give a good understanding of the middle value for both missing models. This will help describe the center of the runtimes compared to the mean and show if the runtimes are skewed in any direction. The descriptive statistics and visual graphs for each simulation condition will provide the information necessary as to which shared parameter method provide the least run time computational challenges.

Chapter III Summary

The methods for this simulation on ILD missing data are setup with the goal to understand the performance, efficiency, and computational intensity between the LTSPMM (Cursio et al., 2019) and the SPLR (Li, X. et al., 2018) missing data models. The data was simulated using the "*simglm*" (LeBeau, 2023) package for four different conditions of assessments and missing data percentages providing a more comprehensive analysis of the shared parameter missing data models. The proposed models guided by the research on marijuana craving and academic motivation by Phillips et al. (2015) adds generalizability to real ILD research. In this simulation, the time dependent predictor academic motivation will be the variable of interest applying both shared parameter missing data models and comparing them with software default missing data model list-wise deletion in terms of raw bias percentage and empirical standard errors.

CHAPTER IV

RESULTS

The purpose of this study was to compare the two joint models used to handle missing data in ecological momentary assessment (EMA) studies and to evaluate their performance under different assessment and missing data scenarios. Joint models simultaneously model the outcome and the missingness process, providing information about the latent trait of responding to prompts in ILD studies. The two joint models that are compared to a poor missing data model like list-wise deletion in this simulation study are the shared parameter logistic regression (SPLR) model proposed by X. Lin et al. (2018) and the Latent Trait Shared Parameter Mixed Model (LTSPMM) proposed by Cursio et al. (2019). The shared parameter logistic regression (SPLR) models the missing process by using a random intercept logistic regression model for the binary missing prompt indicators while the LTSPMM takes into account the missing process using item response theory to model responsiveness to the prompting device as a latent trait. The following research questions guided this study:

Q1 Which model, ILD missing data models LTSPMM (Cursio et al., 2019), SPLR (X. Lin et al., 2018), or the full mixed-effect location random effects model using list-wise deletion, perform better under different combinations of number of prompts (25, 40) and intermittent missingness scenarios (20%, 30%) in terms of raw bias percentage?

Q2 Which model, ILD missing data model LTSPMM (Cursio et al., 2019), SPLR (X. Lin et al., 2018), or the full mixed-effect location random effects model using list-wise deletion perform better under different combinations of number of prompts (25, 40) and missingness scenarios (20%, 30%) in terms of empirical standard errors?

Q3 Which ILD missing data model LTSPMM (Cursio et al., 2019) or the SPLR model (X. Lin et al., 2018), performs more computationally efficient in terms of computational run time?

To address these questions, this study uses the empirical study by Phillips et al. (2015) as a guide to design a simulation that compares the three models under different scenarios of number of prompts (25, 40) and missing conditions (20%, 30%). The performance criteria are parameter estimate raw bias percentage, empirical standard errors, and computation run time. This chapter is organized as follows: first, an overview of the simulation study is provided, followed by the results from the simulation study as well as evaluations of the research questions are provided.

Simulation Overview

The researcher plans to address the research questions by comparing three methods of handling intermittent missingness in intensive longitudinal data: the LTSPMM, the SPLR, and the full mixed-effect location random effects model using list-wise deletion. A simulation study was performed to compare these methods using the statistical software R and the package "simglm" to create the longitudinal data. This simulation study aims to compare a true model with no missing data to three models that handle missing data differently: LTSPMM, SPLR, and full mixed-effect location random effects model using list-wise deletion. The comparison is based on raw bias percentage, empirical standard errors, and computational run time. The simulation varies by the number of assessments (25, 40) and the percentage of intermittent missingness (20%, 30%). For each combination of missing percentages and number of assessments, intensive longitudinal data was generated for 100 participants.

The data is simulated using the "simglm" (LeBeau, 2023) function, which allows for simulating multi-level longitudinal data with ordinal variables and time-dependent covariates.

The function also has an option to generate random intermittent missing data by percentages using an indicator variable. The missing data are considered MCAR and simulated randomly amongst the participants based on an overall missing condition percentage. The missing percentages are set to 20% and 30% to reflect realistic ILD data applications. The data sets are simulated by using predictors from the empirical study with marijuana craving as the outcome variable and determined to be person-specific, and time, motivation, and Rutgers Marijuana Problem Index (RMPI) as the predictor variables. Time represents assessment numbers from 1 to the number of assessments specified (either 25 or 40). Academic motivation is a time dependent variable and is estimated at level 1 as a motivation within predictor and at level 2 as a motivation between predictor. RMPI is a baseline predictor specified to be a person-specific level 2 variable. The regression weight for the mixed-effect model for the intercept was set at 2.41, for time was set at 0.06, for motivation was set at -0.01, and for RMPI was set at 0. Academic motivation was specified to be an ordinal variable with levels ranging from 0 to 10. RMPI was also specified to be an ordinal variable with levels ranging from 0 to 69. Moreover, the variance of the random effect part of the model was set at 10.59.

The LTSPMM (Cursio et al., 2019), the SPLR (X. Lin et al., 2018), and the full mixed-effect location random effects model using list-wise deletion were evaluated on each of the simulated datasets separately. To address the first question, the raw bias percent of the estimated unstandardized regression coefficients for these models were analyzed. The second question is addressed by examining the empirical standard errors of the regression coefficients across one hundred datasets. The third question is analyzed through an assessment of the computational run time and convergence issues of the joint models, which measure the practicality of the models. The performance of the missing data models were evaluated by the parameter estimates associated with time-dependent variable motivation within subjects at level 1 and between subjects at level 2.

Percent Difference Formulas

Percent difference formulas were added to compare how the shared parameter models are performing compared to the missing model that uses list-wise deletion. Equation 4.1 and 4.2 below display an example of these equations for raw bias percent and empirical standard errors

$$Percent\ Difference\ Raw\ Bias = \frac{LD\ Bias\,\% - SP\ Bias\,\%}{LD\ Bias\,\%} \tag{4.1}$$

$$Percent\ Difference\ EmpSE = \frac{LD\ EmpSE - SP\ EmpSE}{LD\ EmpSE}. \tag{4.2}$$

List-wise deletion missing model is a default method in many statistical software packages and is considered a poor missing data method the percent difference calculation will add context to the performance of the shared parameter missing data models.

Research Question 1 Raw Bias Percentage Results

Q1 Which model, ILD missing data models LTSPMM (Cursio et al., 2019), SPLR (X. Lin et al., 2018), or the full mixed-effect location random effects model using list-wise deletion, perform better under different combinations of number of prompts (25, 40) and intermittent missingness scenarios (20%, 30%) in terms of raw bias percentage?

The aim of the first research question is to compare the performance of the LTSPMM, the SPLR, and the full mixed-effect location random effects model using list-wise deletion for handling missing data based on the raw bias percent of the unstandardized regression coefficients estimated from each model. Raw bias percentage represents the percent deviation from the true value of the model with no missing data. The models were evaluated separately based on each simulated data set with different assessments (25, 40) and missing data conditions (20%, 30%). The bias helps quantify how well the missing data methods are estimating the parameters on average with a key property being unbiasedness. The raw bias percent provides information about the accuracy of the coefficient estimates about the true model and the three missing data methods on performance of handling missing data for the simulated data sets.

Table 1 shows the true parameter value estimates calculated as the average of the full mixed-effects location random effects model with no missing data for each assessment number across the simulated datasets for the intercept ($\hat{\beta}_0$) the within-subject motivation ($\hat{\beta}_w$) and between-subject motivation ($\hat{\beta}_B$). After adding the missing data, the estimates from each of the missing data models will be compared to these true values to calculate raw bias percentage and empirical standard errors. The average estimate of within-subject motivation for the 25 assessment were estimated as -0.10 and for the 40 assessment was estimated as -0.12. The average estimate of the between-subject motivation for the 25 assessment were estimated as -0.08 and for the 40 assessment was estimated as -0.1. For more context about the process of this simulation refer to the simulation process in chapter 3 which outlines the data setup and software choices.

Table 1

Simulation Results for True Parameter Value Estimates with no Missing Data

Parameter	25 and 20%	25 and 30%	40 and 20%	40 and 30%
$\boldsymbol{\hat{\beta}_0}$	3.3	3.2	3.4	3.5
$\boldsymbol{\hat{\beta}_B}$	-0.08	-0.08	-0.07	-0.13
$\boldsymbol{\hat{\beta}_w}$	-0.12	-0.07	-0.13	-0.11

Table 2 shows the full table of results of comparing raw bias percentages of the missing data models the LTSPMM, the SPLR, and the full mixed-effect location random effects model using list-wise deletion for different combinations of assessment numbers and missing percentages. In the next few sections, all conditions will be compared by the missing data models for each of the predictors between-subject motivation ($\hat{\beta}_B$) and within-subject motivation ($\hat{\beta}_w$).

Table 2

Comparison of Raw Bias Percentage between List-wise Deletion, SPLR and LTSPMM for Different Number of Prompts and Missing Percentages

		20% Intermittent Missing			30% Intermittent Missing		
Parameter	Number of Assessments	List-wise Deletion %	SPLR %	LTSPMM %	List-wise Deletion %	SPLR %	LTSPMM %
$\hat{\beta}_0$	25	13	7.3	6.9	11	12.7	11.2
	40	8.6	8	13.1	14.2	13.6	13.4
$\hat{\beta}_B$	25	16.1	13.7	13.2	28.9	23.5	17.5
	40	25.6	27	18.6	24.8	23.7	19.2
$\hat{\beta}_w$	25	19.3	13.5	10	40.7	15.3	12.7
	40	40.1	17.2	14.2	33.7	18.4	17.1

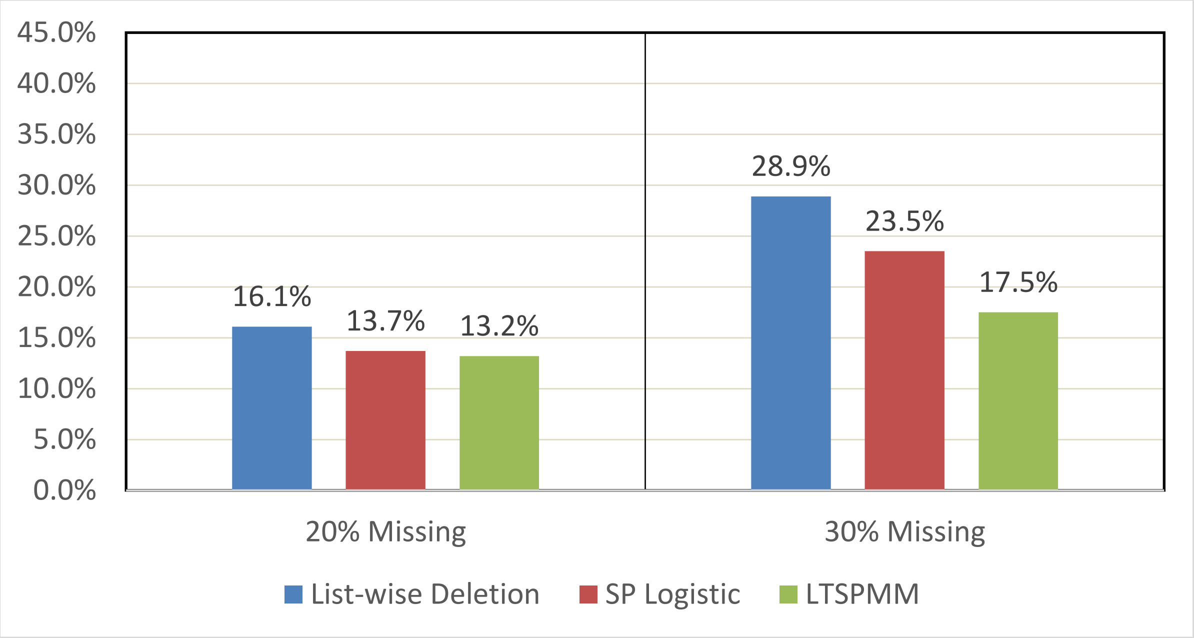
Note. SPLR is Shared Parameter with Logistic Regression and LTSPMM is Latent Trait Shared Parameter Missing Model

25 Assessment Raw Bias Percentage Results

An in depth look at the results for between-subject motivation ($\hat{\beta}_B$) for 25 assessments is displayed below in Figure 3. The full mixed-effect location random effects model using list-wise deletion had a harder time recovering the true estimates with higher raw bias percentages than both shared parameter missing data models for all conditions. Both shared parameter models had similar raw bias percentages at 20% missing performing approximately 15% better compared to list-wise deletion. Increasing the missing to 30% increased the raw bias percentage for all missing data models, however, the LTSPMM displays the best performance amongst the missing data models performing over 39% better than list-wise deletion. SPLR had nearly a 10% increase in raw bias percentage with the increase of missingness from 20% to 30% but still performed over 18% better than list-wise deletion. The LTSPMM missing data model stayed relatively stable with the increase in missingness and outperformed both list-wise deletion and SPLR models for between-subject motivation ($\hat{\beta}_B$) with 30% missing and 25 assessments.

Figure 3

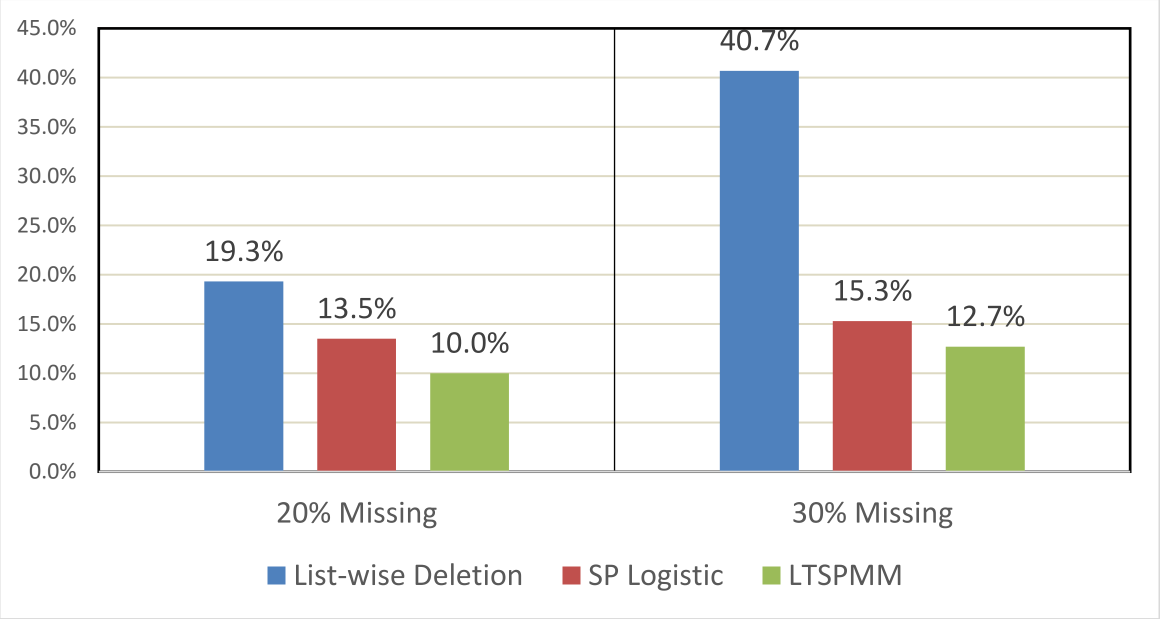
Bar Graph of 25 Prompts Raw Bias Percentage by Missing Data Models for $\hat{\beta}_B$ Between-Subject Motivation



Both shared parameter models had superior raw bias percentages than the list-wise deletion missing data model recovering the predictor within-subject motivation ($\hat{\beta}_w$) at 25 prompts with the results displayed in Figure 4. The SPLR recovered 30% and 62% of the within subject at 20% and 30% missing better than list-wise deletion. The LTSPMM had enhanced performance compared to list-wise deletion with improved bias percentages of 48% at 20% missing and 69% at 30% missing. Both the SPLR and LTSPMM performed competitively in terms of raw bias percentage staying consistent across both missing conditions but the LTSPMM recovered the within-subject motivation marginally better at 25 assessments.

Figure 4

Bar Graph of 25 Prompts Comparing Raw Bias Percentage by Missing Data Models for $\hat{\beta}_W$ Within-Subject Motivation

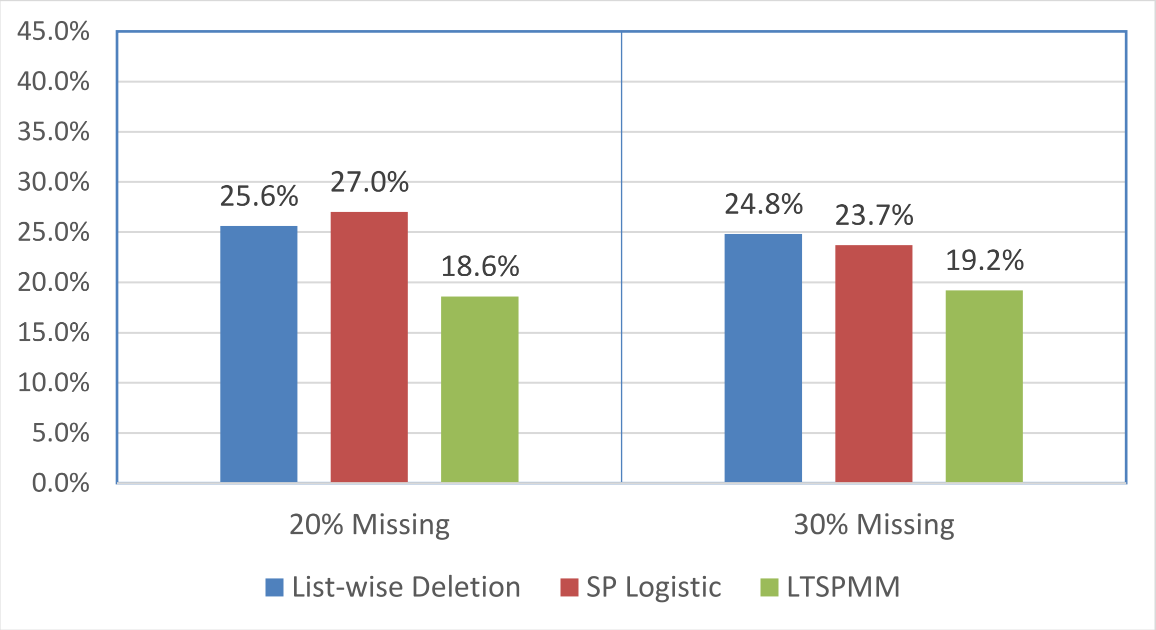


40 Assessment Raw Bias Percentage Results

Between-subject motivation ($\hat{\beta}_B$) for 40 assessments results are presented in Figure 5 with LTSPMM demonstrating superior accuracy in terms of raw bias percentage compared to all of the missing data models. SPLR performed slightly worse than list-wise deletion in terms raw bias percent with 20% missing and slightly better with 30% missing. A surprising result is that the increase in missing to 30% led to both the SPLR and list-wise deletion models to have a slight drop in raw bias percent compared to the 20% indicating that an increase in missingness did not impact the models. Both of these missing data models at 30% missing had similar percentages with 40 assessments as they performed with 25 assessments. LTSPMM performed over 22% better for both missing conditions than list-wise deletion and displayed improvement at capturing the true values compared to the SPLR missing model with decreased percentages of 31% at 20% missing and 19% at 30% missing.

Figure 5

Bar Graph of 40 Prompts Raw Bias Percentages by Missing Data Models for $\hat{\beta}_B$ Between-Subject Motivation

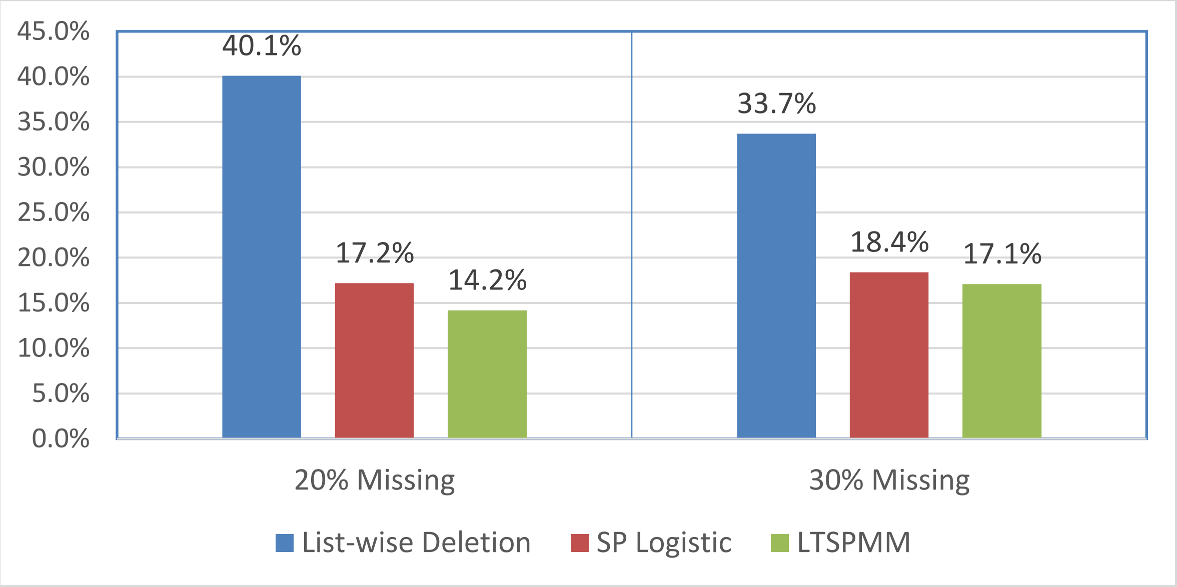


The within-subject motivation ($\hat{\beta}_w$) for 40 prompts revealed an increases in raw bias percentages compared to 25 assessments for all models and is presented in Figure 6 below. The list-wise deletion missing model proved to have troubles capturing the true values with higher raw bias percentages than both shared parameter models for both missing data conditions. At 20% missing data, the LTSPMM performed about 7% better than the SPLR missing data model while both models performed over a 57% improvement than list-wise deletion at this missing percentage. At 30% missing both shared parameter models performed similarly with the LTSPMM demonstrating improved results of 4% compared to the SPLR securing the true values.

Both of the shared parameter models had raw bias percent increases at 40 assessments compared to 25 with the SPLR increasing by approximately 3% and the LTSPMM increasing by over 4% for each condition.

Figure 6

Bar Graph of 40 Assessments Comparing Raw Bias Percentages by Missing Data Models for $\hat{\beta}_W$ *Within-Subject Motivation*



Combined Research Question 1 Results

The LTSPMM performed superior to the other two missing data models with less raw bias percentages for both of the predictors of between-subject motivation ($\hat{\beta}_B$) and within-subject motivation ($\hat{\beta}_w$) across all conditions. Both shared parameter missing data models consistently outperformed list-wise deletion in nearly all conditions displaying the value of the shared parameter missing data models recovering raw bias of time dependent variables in this simulation of ILD longitudinal missing data. The SPLR had a difficult time recovering the between-subject motivation ($\hat{\beta}_B$) for all missing conditions except 25 assessments and 20% missing but remained competitive with the LTSPMM recovering true values for the predictor within-subject motivation ($\hat{\beta}_w$). Therefore, the LTSPMM consistently had lower raw bias percentages for both predictors between-subject motivation ($\hat{\beta}_B$) and within-subject motivation ($\hat{\beta}_w$), this shared parameter model displays an advantage over capturing true values of time-dependent predictors compared to the SPLR missing model for handling missing data. Tables 3 displays the full results of the percent difference calculations for raw bias percent comparing the shared parameter models with list-wise deletion.

Table 3

Percent Difference from Equation 4.1 Results of Raw Bias Percent between List-Wise Deletion Missing Model and Shared Parameter Missing Models

		20% Intermittent Missing		30% Intermittent Missing	
Parameter	Number of Assessments	SPLR %	LTSPMM %	SPLR %	LTSPMM %
$\hat{\beta}_B$	25	14.9	18	19.7	39.4
	40	-5.4	27.3	4.4	22.6
$\hat{\beta}_w$	25	30.1	48.2	62.4	68.8
	40	57.1	64.6	45.4	49.3

Note. SPLR is Shared Parameter with Logistic Regression and LTSPMM is Latent Trait Shared Parameter Missing Model

Research Question 2 Empirical Standard Errors Results

> Q2 Which model, ILD missing data model LTSPMM (Cursio et al., 2019), SPLR (X. Lin et al., 2018), or the full mixed-effect location random effects model using list-wise deletion perform better under different combinations of number of prompts (25, 40) and missingness scenarios (20%, 30%) in terms of empirical standard errors?

The aim of the second research question is to compare the performance of the LTSPMM, SPLR and the full mixed-effect location random effects models using list-wise deletion for handling missing data based on the empirical standard errors of the unstandardized regression coefficients estimated from each model. The empirical standard error is a statistic that estimates the long-run variability of the parameter estimates across multiple simulated data sets generated under similar conditions. It reflects how precise or efficient the estimators are for each model and condition.

Table 4 below shows the results of comparing empirical standard errors between the LTSPMM, SPLR, and the full mixed-effect location random effects model using list-wise deletion for different combinations of prompts and missing percentages. In the next few sections the missing data models will be compared individually by between-subject motivation ($\hat{\beta}_B$) and within-subject motivation ($\hat{\beta}_w$).

Table 4

Comparison of Empirical Standard Errors between List-wise Deletion, SPLR and LTSPMM for Different Number of Assessments and Missing Percentages

		20% Intermittent Missing			30% Intermittent Missing		
Parameter	Number of Assessments	List-wise Deletion	SPLR	LTSPMM	List-wise Deletion	SPLR	LTSPMM
$\hat{\beta}_B$	25	0.70	0.97	0.69	0.66	1.02	0.69
	40	0.73	0.85	0.72	0.66	0.85	0.64
$\hat{\beta}_w$	25	0.48	0.31	0.39	0.68	0.35	0.57
	40	0.50	0.46	0.46	0.68	0.55	0.55

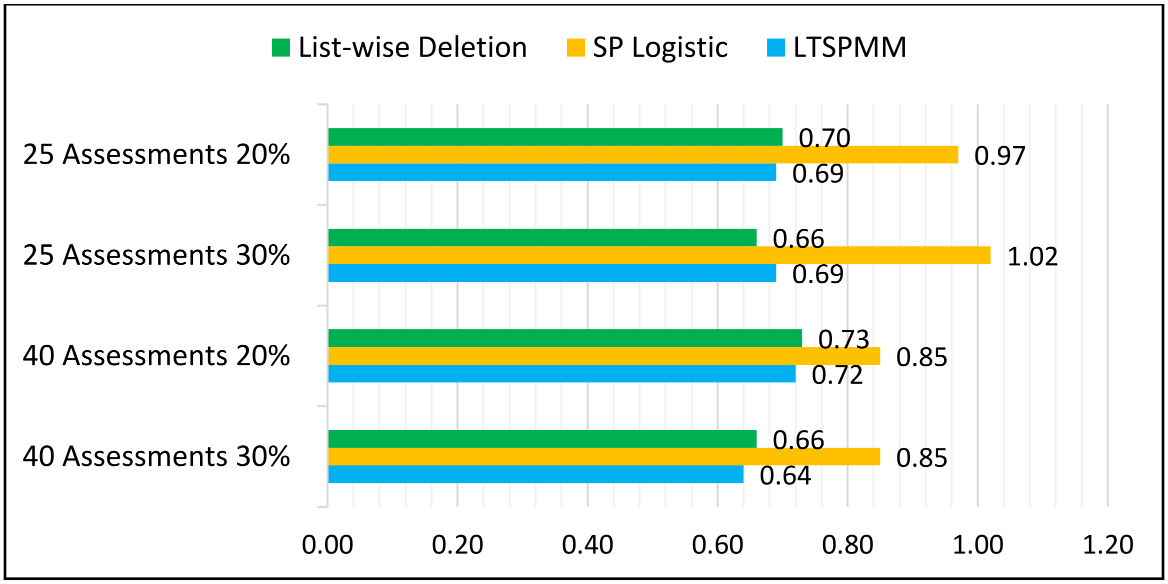
Note. SPLR is Shared Parameter with Logistic Regression and LTSPMM is Latent Trait Shared Parameter Missing Model

Empirical Standard Errors for Between-Subject Motivation ($\hat{\boldsymbol{\beta}}_B$)

The LTSPMM is the best performing and most efficient missing data model in terms of empirical standard errors for between-subject motivation ($\hat{\beta}_B$) narrowly outperforming list-wise deletion in three of the four conditions. Both the LTSPMM and list-wise deletion displayed consistent ranges of empirical standard errors across missing percentages and assessment designs with list-wise performing marginally better when the assessments were 25 with 30% missing. The SPLR missing data model displayed the least precise missing model across all conditions with empirical standard errors increasing for this model at 25 assessments compared to 40. The worst condition for the SPLR model was 25 assessments and 30% missing where the missing model was 54% less precise than list-wise deletion. Figure 7 below displays a visualization of the results for all missing data models.

Figure 7

Bar Graph of all Conditions Comparing Empirical Standard Errors by Missing Data Models for $\hat{\beta}_B$ Between-Subject Motivation

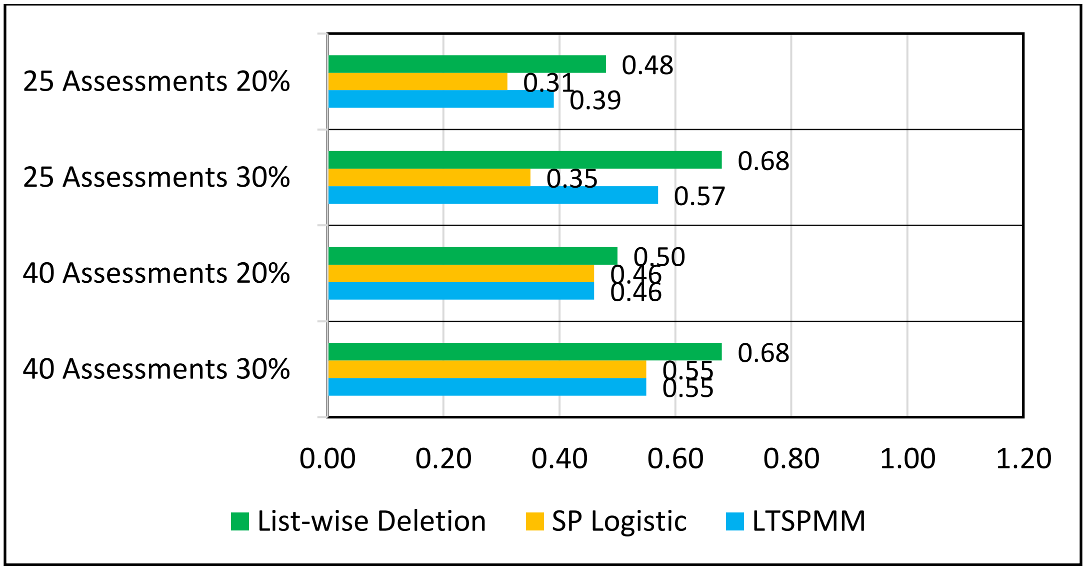


Empirical Standard Errors for Within-Subject Motivation ($\hat{\boldsymbol{\beta}}_w$)

Both shared parameter models performed better and more efficient than list-wise deletion in terms of empirical standard errors for within-subject motivation ($\hat{\beta}_w$) displayed in figure 8. At 40 assessments, the shared parameter models had identical empirical standard errors displaying efficiency that is 8% and 19% better than list-wise deletion across the 20% and 30% missing conditions. At 25 assessments, the SPLR displayed separation from the other models with much lower empirical standard errors where the model performed 35% and 48% better than list-wise deletion across the two missing conditions. The shared parameter models both displayed more efficiency at 25 assessments compared to 40. The LTSPMM was 16% to 19% more efficient than list-wise deletion but 20% and 62% less precise than the SPLR when assessment conditions were set to 25 prompts. Overall, the SPLR displayed the most efficient missing model across all conditions for empirical standard errors for within-subject motivation ($\hat{\beta}_w$).

Figure 8

Bar Graph of all Conditions Comparing Empirical Standard Errors by Missing Data Models for $\hat{\beta}_W$ Within-Subject Motivation



Combined Research Question 2 Results

The LTSPMM displayed the most consistent results across both of the predictors between-subject motivation ($\hat{\beta}_B$) and within-subject motivation ($\hat{\beta}_w$). The SPLR had the worst performance for the predictor between-subject motivation ($\hat{\beta}_B$) but the best overall performance for within-subject motivation ($\hat{\beta}_w$). The condition that this model showed better efficiency than the LTSPMM was at 25 assessments for the within-subject motivation ($\hat{\beta}_w$) predictor. The LTSPMM showed much better efficiency across both predictors for all the assessment and missing conditions displaying enhanced performance of empirical standard errors than the other missing data models. Table 5 presents the full results of percent differences between the shared parameter models and list-wise deletion for all conditions in the simulation.

Table 5

Percent Difference from Equation 4.2 Results of Empirical Standard Errors Between List-Wise Deletion Missing Model and Shared Parameter Missing Models

		20% Intermittent Missing		30% Intermittent Missing	
Parameter	Number of Assessments	SPLR %	LTSPMM %	SPLR %	LTSPMM %
$\hat{\beta}_B$	25	-38.6	1.4	-54.5	-4.5
	40	-16.4	1.4	-28.8	3.0
$\hat{\beta}_w$	25	35.4	18.8	48.5	16.2
	40	8.0	8.0	19.1	19.1

Note. SPLR is Shared Parameter with Logistic Regression and LTSPMM is Latent Trait Shared Parameter Missing Model

Research Question 3 Computational Efficiency Results

Q3 Which ILD missing data model LTSPMM (Cursio et al., 2019) or the SPLR model (X. Lin et al., 2018), performs more computationally efficient in terms of computational run time?

The aim of the third research question is to assess the computational run time of the LTSPMM and the shared parameter location and scale model. Computational run time for the joint missing models provides information on how practical the missing data methods are under different assessment and missingness conditions. For comparison the full mixed-effect location random effects model using list-wise deletion computational run time was instantaneous for each replication taking just minutes to compile the estimates in the simulation. Recall that creators of the two shared parameter missing data models derived and provided code from different statistical software packages. X. Lin et al. (2018) provided code using the rstan package (Stan Development Team, 2022) within the R statistical package (R Core Team, 2021). Cursio et al. (2019) derived code from SAS (SAS Institute Inc, 2013) using the PROC NLMIXED function. To look at the difference between computational runtimes between the LTSPMM and the SPLR missing model averages and medians are calculated and examined. Table 6 below reports the results of runtime comparisons for the LTSPMM and shared parameter location only models for all conditions. The table demonstrates that for all conditions, the SPLR missing data model took substantially more amount of time for each run compared to the LTSPMM. The SPLR model had a median run time range from nearly 23 minutes with 25 assessments at 30% missing to as high as slightly over 42 minutes per run when the assessments rose to 40 with 20% missing. The range for the LTSPMM took much less time ranging from 3 to 8 minutes per run. Neither of the models had convergence issues which might be due to the simplicity of the model. Both shared parameter models display increased computational intensity when missingness is set to 20%

compared to 30%. This demonstrates the computational run time sensitivity these models have to increased sample size. In this simulation, the LTSPMM was significantly less computationally demanding compared to the SPLR for each run of this simulation.

Table 6

Runtime comparison of means and medians for the LTSPMM and SPLR for all conditions (Reported in Minutes)

Conditions	LTSPMM		SPLR	
	Mean	Median	Mean	Median
25 Assessments 20%	209	212	1,610	1,546
25 Assessments 30%	189	189	1,467	1,374
40 Assessments 20%	544	443	2,658	2,538
40 Assessments 30%	479	382	2,490	2,460

Note. SPLR is Shared Parameter with Logistic Regression and LTSPMM is Latent Trait Shared Parameter Missing Model

Chapter IV Summary

In this simulation, the LTSPMM clearly emerged as the best overall performing and most efficient missing data model recovering the time dependent predictor academic motivation for all the conditions. Both shared parameter missing models consistently outperformed list-wise deletion recovering the within-subject motivation ($\hat{\beta}_w$) in terms of raw bias percentage and empirical standard errors. List-wise deletion did have lower empirical standard errors compared to the SPLR for the between-subject motivation ($\hat{\beta}_B$) predictor. However, both shared parameter models displayed value over using statistical software default option list-wise deletion. Finally, the LTSPMM required significantly less computational run time than the SPLR.

CHAPTER V

DISCUSSION AND CONCLUSIONS

The purpose of this study was to compare two joint models used to handle missing data in ecological momentary assessment (EMA) studies and to evaluate their performance under different assessment and missing data scenarios. These joint models were the random intercept logistic regression model proposed by X. Lin et al. (2018) and the model proposed by Cursio et al. (2019) that takes into account the missing process using item response theory to model responsiveness to the prompting device as a latent trait. The study results provide researchers with guidance on the performance of both shared parameter missing data models under missing data conditions that might be observed in real ILD data situations. The designed simulation produced analyses from a few different assessment and missing data scenarios that displays to practitioners the value of using the joint shared parameter missing data models instead of statistical package default missing data method list-wise deletion with improved results in this ILD missing data simulation. Computational intensity still remains an obstacle for researchers on both shared parameter models. Chapter V includes a summary of how study research questions were answered, implications for researchers, limitations, future research, and conclusions.

Research Questions Summary

This study was guided by three research questions. The first research question was formulated to determine the ILD shared parameter missing model that performs better under different combinations of number of prompts and missingness levels in terms of raw bias percentage. It was found that the latent trait shared parameter mixed model (LTSPMM) outperformed the other two missing data models with less raw bias percentages for both of the predictors of between-subject motivation ($\hat{\beta}_B$) and within-subject motivation ($\hat{\beta}_w$) across all conditions. The shared parameter with logistic regression (SPLR) missing model displayed value as an EMA missing data model with improved results compared to list-wise deletion. The second research question was formulated to determine the ILD shared parameter missing model that performs better under different combinations of number of prompts and missingness levels in terms of empirical standard errors. Results of the simulation demonstrated that the SPLR outperformed the LTSPMM and list-wise for within-subject motivation ($\hat{\beta}_w$). However, the LTSPMM performed significantly better than the SPLR for the predictor between-subject motivation ($\hat{\beta}_B$) and displayed the most consistency across all conditions making it the better overall model for empirical standard errors. The third research question was formulated to determine the ILD shared parameter missing model that performs better in terms of computational run time. The results revealed that for all conditions, the SPLR model took a substantially more amount of time to run compared to LTSPMM. These results are somewhat surprising as Cursio et al. (2019) outlined computational intensity with their missing shared parameter model.

Implications for Researchers

This study reinforces the use of both the LTSPMM and SPLR missing data models to inform missingness in ILD research studies with the focus on the performance of recovering and stabilizing time-dependent variables. The LTSPMM emerged as the superior overall missing data model in this simulation and is recommended from this research in all ILD missing data situations. The impact of raw bias on the shared parameter missing data models increased as the number of assessments went from 25 to 40 and missing data changed from 20% to 30% in the study. Researchers should expect higher amounts of raw bias with additional assessments. Empirical standard errors displayed different results between the within-subject and between-subject predictors amongst the conditions. The within-subject predictor displayed clear trends of increased empirical standard errors as the intermittent percentage increased from 20% to 30% and assessments changed from 25 to 40 indicating that the models become less precise with additional missing data and more assessments. Empirical standard errors of the between-subjects predictor displayed the opposite trend as the models became more precise as assessments increased from 25 to 40.

The major drawback for researchers of the joint shared parameter missing data models is the computational intensity. In this simulation, the LTSPMM required significantly less computational run time than the SPLR with run time results for both missing data models much better than expected. Simplicity of the model might be a key reason as Cursio et al. (2019) outlined convergence issues with a large numbers of parameters needing at least 3 hours to converge for all one-parameter models. In this simulation, there were no convergence issues and run time of the LTSPMM ranged from 3 to 8 minutes with the SPLR ranging from 23 to 42 minutes. Computational intensity increased significantly from 25 assessments to 40 as the percent difference in computational run time increased by 68% for the LTSPMM and 57% for the SPLR. Researchers should be encouraged by these results but also aware that adding parameters, assessments, and sample size will intensify the computational run time of the joint shared parameter missing data models.

The SPLR missing model consistently outperformed list-wise deletion in terms of raw bias percentage for both the between-subject and within-subject predictors suggesting its use over poor default missing data methods. While the LTSPMM did display better overall raw bias percentages, the SPLR remained competitive across all conditions. The major difference between the models was the performance of the SPLR in recovering the between-subject motivation predictor specifically which resulted in estimating less precise for all conditions in terms of empirical standard errors. In contrast, the SPLR performed more efficient than the LTSPMM for the within-subject predictor on empirical standard errors at both missing conditions and 25 assessments while performing the same at 40 assessments. These results are exciting as the SPLR features a shared parameter scale model that adds information about the missingness of the within-subject variation and the primary outcome (equation 2.56). Researchers are encouraged to implement the SPLR when research questions are related to within-subject predictors as the SPLR is an excellent option to validate and inform the effect of missingness on the outcome.

The generalizability of this simulation is useful to ILD studies with missing data outside of the rates chosen. The performance of the shared parameters displayed clear trends of improved raw bias and empirical standard errors of the within-subject estimates compared to list-wise deletion for all prompt and missing conditions. The shared parameter models performed at least 30% better than list-wise deletion recovering the true estimates and at least 8% improved for empirical standard errors of the within-subject predictor for every condition. Even if researchers come across missing data ranges outside of the conditions in this study the trends indicate that the shared parameter missing data models should be implemented for better performance of recovering time dependent predictors. This simulation studied the performance of the recent ILD missing data models but researchers should not forget that the true purpose of the models is to provide informative missingness. The shared parameters implement the latent trait of participants responding to the prompting device, which is modeled jointly with a mixed model for bivariate longitudinal outcomes. Even when missing data is outside the conditions ranges in this study, it is recommended that researchers apply either the LTSPMM or the SPLR for improved performance and informative missingness in all ILD studies.

Publication Bias

This study applied compliance rates from meta-analysis on ILD studies as a tool to help guide the study conditions for missing data. Publication bias may occur in meta-analysis that distorts any attempt to derive valid estimates by skewing compliance towards higher rates (Thornton & Lee, 2000). One concern to consider is the publication bias of studies with lower compliance rates, which researchers may not submit or accept for publication limiting the number of studies in the meta-analysis. The authors of the meta-analysis found inconsistencies of reported compliance for published articles. Numerous studies could not be included in our analyses due to an absence of compliance data reported (Jones et al., 2019; Williams et al., 2021). It is possible that researchers did not report poor compliance rates inflating the rates found in the meta-analysis.

Another concern is how researchers handled subjects with poor compliance within their studies. Studies included in meta-analyses privilege best compliers through exclusion of participants not meeting criteria for valid EMA data or compliance thresholds (determined a priori or posteriori) (Williams et al., 2021). While Jones et al. (2019) admitted that compliance rates are likely inflated because some participants who did not reach a specific rate of compliance (our supporting information analyses demonstrated that approximately 6% of participants were excluded from studies due to poor compliance).

A last concern to mention is the publication bias related to sponsorship. Thornton and Lee (2000) stress that a study's source of funding may also unduly influence the probability of subsequent publication of the results. Researchers need funding to conduct costly ILD studies to cover the expenses of technology and financial incentives for participants. Shiffman et al. (2008) warn that the high fixed costs of technological solutions can make it very hard to initiate small studies or pilot studies. Small studies may not even submit for publication or unlikely to be published due to smaller sample sizes, poorer compliance rates, or both. Thus, small non-funded studies are unlikely to be included in meta-analyses increasing the chances of publication bias.

Study Limitations

The results and conclusions of this study are limited to only four different model conditions that were simulated in this study. Due to the computational intensity of the shared parameter missing models the conditions had to be limited to ensure completion of the simulation study, but many more models could have been selected. ILD researchers choose many different assessments designs depending on the topic of interest but only 25 and 40 assessment conditions were chosen based on an empirical study but fewer or more assessments could have easily been chosen to study. The sample size was the same across all conditions at n=100 but a wide range of sample sizes could have been preferred and may influence the shared parameter missing data models. The specification of an uncomplicated model to understand missing data might not fit real research data situations that incorporate many variable types with diverse distributions making the missing data even more complex. The linear mixed effects random intercept location only model was chosen for its simplicity in understanding raw bias and empirical standard errors but adding a random slope might be warranted in some longitudinal data situations.

Missing data was randomly intermittent at twenty and thirty percent based on the compliance rate findings from meta-analyses. The compliance rates chosen could be inflated due to low compliance rates not being published, compliance not being reported, exclusion of non-compliers, and lack of sponsorship for small studies. Missing data that is MNAR was not chosen to implement for this study and could possibly be the missing mechanism for ILD studies. Both X. Lin et al. (2018) and Cursio et al. (2019) studied their missing data models after implementing MNAR missing data reporting positive results. However, in this simulation, the missing data was introduced intermittently at the within participant level MCAR that did not account for any missing data patterns that may emerge in real ILD missing data situations.

Future Research

Due to the computational nature of this ILD missing data simulation study, the number of conditions were limited. Future studies could expand on the conditions adding a more thorough understanding of the shared parameter missing data models. For example, a different number of assessments and increased missing data percentages could provide guidance of a more diverse range of ILD missing data situations. Recall that the sample size for this study remained constant at 100 participants across all conditions in the simulation. Studies with smaller sample sizes is an area that needs to be explored. With the costly nature of conducting ILD studies, researchers without funding might have less participants to study with lower compliance rates and understanding missing data in these situations could help improve their analyses. As discussed in the sample size section in chapter 3, there is a wide range in sample sizes amongst ILD researchers and different participant ranges could influence the shared parameter missing data models.

The SPLR (X. Lin et al., 2018) included a shared parameter to the random subject scale (variance) as an extension to location model of the LTSPMM (Cursio et al., 2019). This allows the random location and scale to influence missingness. Within subject variation of the primary outcome is a strong component of ILD research and connecting unstable subjects with higher variation may influence subject missingness. The LTSPMM missing model was studied as a random location linear mixed effect model and has not added the random scale effect. Adding the random subject scale effect to the LTSPMM seems to be a natural extension to the missing model.

In this study, the LTSPMM implemented a one-parameter (1PL) logistic IRT model that represents the latent trait of "responsiveness" and corresponds to how each participant responds to the prompting device. The choice to use the 1PL model was due to the computational intensity of the 2PL described by Cursio et al. (2019). Recall that the two-parameter (2PL) IRT logistic model that allows for each the discrimination parameters a_t to have a unique slope for each time-bin accounting for more information in the model. The results of their study found that 2PL model outperformed the 1PL LTSPMM in terms of bias and standard error of the estimated regression coefficient of the latent trait θ_i. In this study, the 1PL LTSPMM had surprisingly less computational intensity than expected and resulted in the best performing ILD missing data model. Allowing the discrimination parameter to vary across time-bins adds more context for researchers to understand their missing data situation. Thus, more exploration on the 2PL LTSPMM missing model feels like a valuable next step in ILD missing data research continuum.

The computational demands of the shared parameter missing data models limited the number of conditions that could be studied in this simulation and remains a drawback for researching the models in large data situations. The sensitivity of the computational intensity was displayed in the computational efficiency results section in chapter four that run time increased for both shared parameter missing data models when missingness was set to 20% compared to 30%, which only added 250 or 400 data points to the estimated datasets. In recent research, resampling methods via representative points has become a popular artificial intelligence technique in order to handle large data sets reducing computational run time. The idea is to take a small sample from the full data set confirming it closely approximates the distribution of the larger sample making the small sample generalizable to the larger data set. Recent research on the topic displays multiple methods to reseample the data like Neighborhood-Based Cross Fitting (NBCF) (Agboola, 2022), Double Super Learner (DSL) (Alanazi, 2022) and the Support Points Sample Splitting (SPSS). Gao et al. (2022) introduced a method with missing data called MissDAG. Reducing the computational burden of the shared parameter missing data models on large data sets could advance literature and research on the these models making them more applicable in future EMA studies.

Final Conclusions

The benefits of this simulation study on missing ILD data have both practical and theoretical application. The performance and usefulness of the shared parameter models over statistical software default missing data method list-wise deletion was repeatedly displayed throughout the simulation adding a theoretical benefit for both statisticians and researchers to implement these models to handle ILD missing data. Computational run time was more practical in this study with models compiling estimates in minutes instead of hours giving researcher’s the ability to understand if missing data are associated with their study outcomes in a shorter expected run time. Comparing the shared parameter models was a logical step to advance the literature on missing data for ILD studies. The goal was to compare the two shared parameter models on the same dataset under various conditions of intermittent missing data and assessments to understand their performance in terms of raw bias, empirical standard errors and computational run time. The missing model from Cursio et al. (2019) LTSPMM emerged as the top performing missing data model in this simulation but the SPLR model by X. Lin et al. (2018) still showed promise and could be very useful if research questions involve within subject variance. ILD researchers are implementing an assortment of different predictor types, assessment designs and sample sizes and further research is encouraged to understand more about the inherent complex missing data in these studies. The results of this study offers researchers more information and confidence regarding implementation of either the LTSPMM or the SPLR missing data methods to understand missing data in their ILD studies.

REFERENCES

Agboola, O. D., (2022). An Efficient Computational Method for Causal Inference in High-Dimensional Data: Neighborhood-Based Cross Fitting, *University of Northern Colorado*. 845. https://digscholarship.unco.edu/dissertations/845

Alanazi, S. S., (2022). An Ensemble Machine Learning Approach To Causal Inference in High-Dimensional Settings. *University of Northern Colorado*. 919. https://digscholarship.unco.edu/dissertations/919

Albert, P. S., & Follmann, D. A. (2009). Shared-parameter models. *Longitudinal data analysis*, 433-452.

Anderson, T.W. (1957). Maximum likelihood estimates for a multivariate normal distribution when some observations are missing. *Journal of the American Statistical Association*, *52*, 200–203. doi:10.1080/01621459.1957.10501379.

Aouar, L., (2023). An Adaptive Deep Learning for Causal Inference Based on Support Points With High-Dimensional Data. *University of Northern Colorado*. 1023. https://digscholarship.unco.edu/dissertations/1023

Baraldi, A. N., & Enders, C. K. (2010) An introduction to modern missing data analyses, *Journal of School Psychology*, Volume 48, Issue 1, Pages 5-37, ISSN 0022-4405, https://doi.org/10.1016/j.jsp.2009.10.001.

Black, A. C., Harel, O., & Matthews, G. (2011) Techniques for modeling intensive longitudinal data with missing values. In T. S. Conner & M. Mehl (Eds.). Handbook of research methods for modeling daily life (pp. 339-356). New York: Guilford Press.

Bock, R. D., & Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: Application of an EM algorithm. *Psychometrika, 46*(4), 443-4459. https://doi.org/10.1007/BF02293801

Bock, R. D. (1989). *Multilevel analysis of educational data.* New York: Academic Press.

Bodner, T. E. (2006). Missing data: prevalence and reporting practices. *Psychological Reports*, *99*, 675–680. doi:10.2466/PR0.99.7.675-680.

Bolger, N., Davis, A., & Rafaeli, E. (2003). Diary methods: Capturing life as it is lived. *Annual. Review of Psychology*. 54, 579–616.

Buckner, J. D., Crosby, R. D., Silgado, J.,Wonderlich, S. A., & Schmidt, N. B. (2012). Immediate antecedents of marijuana use: An analysis from ecological momentary assessment. Journal of Behavior Therapy and Experimental Psychiatry, 43(1), 647–655. http://dx.doi.org/10.1016/j.jbtep.2011.09.010.

Casella, G., & George, E. (1992). Explaining the Gibbs sampler. *American Statistics*, *46*(3), 167–174.

Cursio, J. F., Mermelstein, R. J., & Hedeker, D. (2019). Latent trait shared-parameter mixed models for missing ecological momentary assessment data. *Statistics in medicine, 38*, 4, 660-673.

De Silva, A. P., Moreno-Betancur M., De Livera A. M., Lee K. J., Sompson J. A. (2017) A comparison of multiple imputation methods for handling missing values in longitudinal data in presence of a time-varying covariate with a non-linear association with time: a simulation study. *BMC medical research methodology*, *17,* (114).

Demirtas, H., & Schafer, J. L. (2003). On the performance of random-coefficient patternmixture models for non-ignorable dropout. *Statistics in Medicine 22*, 2533–2575.

de Vries, L. P., Baselmans, B. M. L., & Bartels, M. (2021). Smartphone-Based Ecological Momentary Assessment of Well-Being: A Systematic Review and Recommendations for Future Studies. *Journal of happiness studies, 22*(5), 2361–2408. https://doi.org/10.1007/s10902-020-00324-7

deVries, M. W. (Ed.). (1992). The experience of psychopathology: Investigating mental disorders in their natural settings. Cambridge University Press. https://doi.org/10.1017/CBO9780511663246

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

Donders, A. R. T., Van Der Heijden, G. J., Stijnen, T., & Moons, K. G. (2006). A gentle introduction to imputation of missing values. *Journal of clinical epidemiology, 59*(10), 1087-1091.

Dong, Y., & Peng, C. Y. J. (2013). Principled missing data methods for researchers. *Springer Plus*, *2*(1), 222.

Enders, C. K. (2001) A Primer on Maximum Likelihood Algorithms Available for Use With Missing Data. *Structure Equation Modeling*, *8*(1):128–141. doi:10.1207/S15328007SEM0801_7

Enders, C. K., & Bandalos, D. L. (2001). The relative performance of full information maximum likelihood estimation for missing data in structural equation models. *Structural Equation Modeling*, *8*, 430–457. doi:10.1207/S15328007SEM0803_5.

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

Enders, C. K., Mistler, S. A., & Keller, B. T. (2016). Multilevel Multiple Imputation: A Review and Evaluation of Joint Modeling and Chained Equations Imputation. *Psychological Methods* 21 (2). 222–40.

Enders, C. K., Keller, B. T., & Levy, R. (2018). A fully conditional specification approach to multilevel imputation of categorical and continuous variables. *Psychological Methods*, *23*(2), 298-317. https://doi.org/10.1037/met0000148

Fitzmaurice, G. M., Laird, N. M. & Shneyer, L. (2001). An alternative parameterization of the general linear mixture model for longitudinal data with non-ignorable drop-outs. Statist. Med., 20: 1009-1021. doi:10.1002/sim.718

Gao, E., Ng, I., Gong, M., Shen, L., Huang, W., Liu, T., & Bondell, H. (2022). Missdag: Causal discovery in the presence of missing data with continuous additive noise models. *Advances in Neural Information Processing Systems*, *35*, 5024-5038.

Glynn, R. J., Laird, N. M., & Rubin, D. B. (1986). Selection modelling versus mixture modelling with nonignorable nonresponse. In H. Wainer (ed.), *Drawing Inferences from Self-Selected Samples*, pp. 115–142. New York: Springer

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

Graham, J. W. (2009). Missing data analysis: Making it work in the real world. *Annual Review of Psychology*, *60*, 549–576.

Goldstein, H., Carpenter, J., Kenward, M. G., & Levin, K. A. (2009). Multilevel models with multivariate mixed response types. *Statistical Modelling*, *9*(3), 173–197. https://doi.org/10.1177/1471082X0800900301

Greenland, S., & Finkle, W. D. (1995). A critical look at methods for handling missing covariates in epidemiologic regression analyses. *American journal of epidemiology*, *142*(12), 1255-1264.

GSMA. (2019). The Mobile Economy 2019 https://www.gsma.com/r/mobileeconomy/

Grund, S., Lüdtke, O., & Robitzsch, A. (2018). Multiple Imputation of Missing Data at Level 2: A Comparison of Fully Conditional and Joint Modeling in Multilevel Designs. *Journal of Educational and Behavioral Statistics.* doi.org/10.3102/1076998617738087.

Hayati Rezvan, P., Lee, K. J., & Simpson, J. A. (2015). The rise of multiple imputation: a review of the reporting and implementation of the method in medical research. *BMC Medical Research Methodology*, *15*(1).

Heckman, 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.

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

Hedeker, D., Mermelstein, R. J., & Demirtas, H. (2008). An application of a mixed-effects location scale model for analysis of Ecological Momentary Assessment (EMA) data. *Biometrics*, *64*(2), 627–634. https://doi.org/10.1111/j.1541-0420.2007.00924

Hektner, J. M., Schmidt, J. A., & Csikszentmihalyi, M. (2007). Experience sampling method: Measuring the quality of everyday life. Sage Publications, Inc.

Hickey, G. L., Philipson, P., Jorgensen, A., & Kolamunnage-Dona, R. (2016). Joint modelling of time-to-event and multivariate longitudinal outcomes: recent developments and issues. *BMC medical research methodology*, *16*(1), 117. https://doi.org/10.1186/s12874-016-0212-5

Honaker, J., King, G., & Blackwell, M. (2011). Amelia II: A program for missing data. *Journal of Statistical Software, 45* (7), 1–47.

Hogan, J. W., & Laird, N. M. (1997). Mixture models for the joint distribution of repeated measures and event times. *Statistics in medicine*, *16*(3), 239-257.

Hughes, R. A., White, I. R., Seaman, S. R., Carpenter, J. R., Tilling, K., & Sterne, J. A. C. (2014). Joint modeling rationale for chained equations. *BMC Medical Research Methodology*, *14,* 1–10. http://dx.doi.org/10.1186/1471-2288-14-28

Huque, M. H., Carlin, J. B., Simpson, J. A. & Lee, K. J. (2018). A comparison of multiple imputation methods for missing data in longitudinal studies. *BMC Medical Research Methodology*, *18* (1), https://doi.org/10.1186/s12874-018-0615-6.

Ibrahim, J. G., Chen, M. H., & Lipsitz, S. R. (2001). Missing responses in generalised linear mixed models when the missing data mechanism is nonignorable. *Biometrika*, *88*(2), 551-564.

Jahng, S., Wood, P. K., & Trull, T. J. (2008). Analysis of affective instability in ecological momentary assessment: Indices using successive difference and group comparison via multilevel modeling. *Psychological Methods, 13,* 354–375.

Jelicic, H., Phelps, E., & Lerner, R. M. (2009). Use of missing data methods in longitudinal studies: The persistence of bad practices in developmental psychology. *Developmental Psychology*, *45,* 1195–1199. doi:10.1037/a0015665.x

Ji, L., Chow, S. M., Schermerhorn, A. C., Jacobson, N. C., & Cummings, E. M. (2018). Handling Missing Data in the Modeling of Intensive Longitudinal Data. *Structural* *equation modeling : a multidisciplinary journal*, *25*(5), 715–736. https://doi.org/10.1080/10705511.2017.1417046

Jones, A., Remmerswaal, D., Verveer, I., Robinson, E., Franken, I. H. A., Wen, C. K. F., & Field, M. (2019) Compliance with ecological momentary assessment protocols in substance users: a meta-analysis, *Addiction*. 114, 609– 619. doi: https://doi.org/10.1111/add.14503.

Kunkel, D., & Kaizar, E. E. (2017). "A Comparison of Existing Methods for Multiple Imputation in Individual Participant Data Meta-Analysis." *Statistics in Medicine* 36 (22). 3507–32

Laird, N. M., & Ware, J. H. (1982). Random-effects models for longitudinal data. *Biometrics*, *38*, 963–974.

Lang, K. M., & Little, T. D. (2018). Principled Missing Data Treatments. *Prevention science : the official journal of the Society for Prevention Research*, *19*(3), 284–294. https://doi.org/10.1007/s11121-016-0644-5

Larson, R., & Csikszentmihalyi, M. (1983). The Experience Sampling Method. *New Directions for Methodology of Social & Behavioral Science*, *15*, 41–56.

LeBeau, B. (2023). *simglm: Simulate Models Based on the Generalized Linear Model*. R package version 0.9.5, https://github.com/lebebr01/simglm.

Liao, Y., Skelton, K., Dunton, G., & Bruening, M. (2016). A systematic review of methods and procedures used in ecological momentary assessments of diet and physical activity research in youth: An adapted STROBE checklist for reporting EMA Studies (CREMAS). *Journal of Medical Internet Research*, *18*(6), [e151]. https://doi.org/10.2196/jmir.4954

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.

Lin, X., Mermelstein, R., & Hedeker, D. (2018). A shared parameter location scale mixed effect model for EMA data subject to informative missing. *Health Services and Outcomes Research Methodology, 18,* 227-243.

Little, R. J. (1988). A test of missing completely at random for multivariate data with missing values. *Journal of the American Statistical Association*, *83*(404), 1198–1202.

Little, R. J. (1993). Pattern-mixture models for multivariate incomplete data. *Journal of the American Statistical Association*, *88*(421), 125-134.

Little, R. J. (1994). A class of pattern-mixture models for normal incomplete data. *Biometrika*, *81*(3), 471-483.

Little, R. J. (1995). Modeling the drop-out mechanism in repeated-measures studies. *Journal of the American Statistical Association*, *90*(431), 1112-1121.

Little, R.J.A. and Rubin, D.B. (2002) Statistical Analysis with Missing Data. 2nd Ed., Wiley Interscience, New York. https://doi.org/10.1002/9781119013563

Liu, L. (2008). A model for incomplete longitudinal multivariate ordinal data. Statistics in *Medicine*, *27,* 6299–6309.

Liu, L., & Hedeker, D. (2006). A Mixed-Effects Regression Model for Longitudinal Multivariate Ordinal Data. *Biometrics*, *62*(1), 261-268.

Lord, F. M., & Novik, M. R. (1968). *Statistical theories of mental test* scores. Reading, Mass Addison- Wesley Pub. Co.

McLean, D. C., Nakamura, J., & Csikszentmihalyi, M. (2017). Explaining System Missing: Missing Data and Experience Sampling Method. *Social Psychological and Personality Science*, *8*(4), 434–441. https://doi.org/10.1177/1948550617708015

Messiah, A., Grondin, O., & Encrenaz, G. (2011). Factors associated with missing data in an experience sampling investigation of substance use determinants. *Drug and alcohol dependence*, *114*(2-3), 153–158. https://doi.org/10.1016/j.drugalcdep.2010.09.016

Mistler, S. A., & Enders, C. K. (2017). "A Comparison of Joint Model and Fully Conditional Specification Imputation for Multilevel Missing Data." *Journal of Educational and Behavioral Statistics* 42 (4): 432–66.

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

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

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

Molenberghs, G. & Verbeke, G. (2001). A review on linear mixed models for longitudinal data, possibly subject to dropout. *Statistical Modelling, 1,* 235-269.

Nakai, M. & Ke, W. (2011). Review of the Methods for Handling Missing Data in Longitudinal Data Analysis. *Journal of Math Analysis*. *5*. 1-13.

Nevalainen, J., Kenward, M. G., & Virtanen, S. M. (2009). Missing values in longitudinal dietary data: a multiple imputation approach based on a fully conditional specification. *Statistics in medicine, 28 29,* 3657-69.

Olinsky, A., Chen, S., & Harlow, L. (2003). The comparative efficacy of imputation methods for missing data in structural equation modeling. *European Journal of Operational Research*, *151*, 53–79.

Peng, C. Y., Harwell, M. R., Liou, S. M., & Ehman, L. H. (2006). Advances in missing data methods and implications for educational research. In S. S. Sawilowsky (Ed.), *Real Data Analysis* (pp. 31-78).

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

Phillips, K. T., Phillips, M. M., Lalonde, T. L., & Tormohlen, K. N. (2015). Marijuana use, craving, and academic motivation and performance among college students: An in-the-moment study. *Addictive behaviors*, *47*, 42–47. https://doi.org/10.1016/j.addbeh.2015.03.020

Quartagno, M., & Carpenter, J. R. (2016) Multiple imputation for IPD meta-analysis: allowing for heterogeneity and studies with missing covariates. *Statist. Med.*, 35: 2938– 2954. doi: 10.1002/sim.6837

R Core Team (2021). R: A language and environment for statistical ## computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Rasch, G. (1960). Studies in mathematical psychology: I. Probabilistic models for some intelligence and attainment tests.

Resche-Rigon, M., & White, I. R. (2018). Multiple imputation by chained equations for systematically and sporadically missing multilevel data. *Statistical Methods in Medical Research*, *27*(6), 1634–1649. https://doi.org/10.1177/0962280216666564

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

Roy, J., & Daniels, M. J. (2008). A general class of pattern mixture models for nonignorable dropout with many possible dropout times. *Biometrics*, *61*, 538–545.

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

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

Rubin, D. B., & Schafer, J. L. (1990). *Efficiently creating multiple imputations for incomplete multivariate normal data.* Paper presented at the Proceedings of the Statistical Computing Section of the American Statistical Association.

SAS Institute Inc (2013). **SAS**/ACCESS® 9.4 Interface to ADABAS: Reference. Cary, NC, SAS Publishing.

Schafer, J. L. (1997). *Analysis of incomplete multivariate data*. New York: Chapman & Hall/CRC.

Schafer, J. L. (2001). Multiple imputation with PAN. In A. G. Sayer & L.M. Collins (Eds.), *New methods for the analysis of change* (pp. 355–377). Washington, DC: American Psychological Association. http://dx.doi.org/10.1037/10409-012

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

Schafer, J. L., & Yucel, R. M. (2002). Computational strategies for multivariate linear mixed effects models with missing data. *Journal of Computational and Graphical Statistics, 11,* 437–457. http://dx.doi.org/10.1198/106186002760180608

Shrier, L. A., Walls, C. E., Kendall, A. D., & Blood, E. A. (2012). The context of desire to use marijuana: Momentary assessment of young people who frequently use marijuana.

Psychology of Addictive Behaviors, *26*(4), 821–829. http://dx.doi.org/10.1037/a0029197.

Shiffman S, Stone, A. A., & Hufford, M. R. (2008). Ecological momentary assessment. *Annual Review of Clinical Psychology*. *4*, 1–32.

Shih, W. J., Quan, H., & Chang, M. N. (1994). Estimation of the mean when data contain non-ignorable missing values from a random effects model. *Statistics and Probability Letters, 19*, 249–257.

Silvia, P. J., Kwapil, T. R., Eddington, K. M., & Brown, L. H. (2013). Missed beeps and missing data: Dispositional and situational predictors of nonresponse in experience sampling research. *Social Science Computer Review, 31*(4), 471-481. https://doi.org/10.1177/0894439313479902

Sokolovsky, A. W., Mermelstein, R. J., & Hedeker, D. (2014). Factors predicting compliance to ecological momentary assessment among adolescent smokers. *Nicotine & tobacco research : official journal of the Society for Research on Nicotine and Tobacco*, *16*(3), 351–358. https://doi.org/10.1093/ntr/ntt154

Stan Development Team (2022). RStan: the R interface to Stan. R package version 2.21.5. https://mc-stan.org/.

Stone, A. A., & Shiffman, S. (1994). Ecological momentary assessment (EMA) in behavorial medicine. *Annals of Behavioral Medicine*.

Stone, A. A., & Shiffman, S. (2002). Capturing Momentary, Self-Report Data: A Proposal for Reporting Guidelines. *Annals of behavioral medicine: a publication of the Society of Behavioral Medicine*. *24*. 236-43. 10.1207/S15324796ABM2403_09.

Stone, A. A., Shiffman, S., & Atienza, A. (2007). The Science of Real-Time Data Capture: Self-Reports in Health Research.

Thijs, H., Molenberghs, G., Michiels, B., Verbeke, G., & Curran, D. (2002). Strategies to fit pattern-mixture models. *Biostatistics, 3*(2), 245-265.

Thornton, A., & Lee, P. (2000). Publication bias in meta-analysis: its causes and consequences. *Journal of clinical epidemiology*, *53*(2), 207–216. https://doi.org/10.1016/s0895-4356(99)00161-4

van Buuren, S. (2011). Multiple imputation of multilevel data. In J. K. Roberts & J. J. Hox (Eds.), *The handbook of advanced multilevel analysis* (pp. 173–196). New York, NY: Routledge.

van Buuren, S. (2018). Flexible Imputation of Missing Data, Second Edition. New York: Chapman and Hall/CRC, https://doi.org/10.1201/9780429492259

van Buuren, S., Brand, J. P. L., Groothuis-Oudshoorn, C. G. M., & Rubin, D. B. (2006). Fully conditional specification in multivariate imputation. *Journal of statistical computation and simulation*, *76*(12), 1049-1064.

van Buuren, S., & Groothuis-Oudshoorn, K. (2011). mice: Multivariate imputation by chained equations in r. *Journal of Statistical Software*, *45* (3), 1–67. Retrieved from http://www.jstatsoft.org/v45/i03/

van Ginkel, J. R., Linting, M., Rippe, R. C., & van der Voort, A. (2020). Rebutting existing misconceptions about multiple imputation as a method for handling missing data. *Journal of personality assessment*, *102*(3), 297-308.

Walls, T. A., & Schafer, J. L. (Eds.). (2006). Models for intensive longitudinal data. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195173444.001.0001

Wen, C., Schneider, S., Stone, A. A., & Spruijt-Metz, D. (2017). Compliance With Mobile Ecological Momentary Assessment Protocols in Children and Adolescents: A Systematic

Review and Meta-Analysis. *Journal of medical Internet research*, *19*(4), e132. https://doi.org/10.2196/jmir.6641

White, H. R., Labouvie, E. W., & Papadaratsakis, V. (2005). Changes in substance use during the transition to adulthood: A comparison of college students and their non-college age peers. *Journal of Drug Issues*, *35*(2), 281–306. http://dx.doi.org/10.1177/ 002204260503500204.

Wilkinson, L. (1999). Statistical methods in psychology journals: guidelines and explanations. *American Psychologist*, *54*, 594–604. doi:10.1037//0003-066X.54.8.594.

Williams, M. T., Lewthwaite, H., Fraysse, F., Gajewska, A., Ignatavicius, J., & Ferrar, K. (2021). Compliance with mobile ecological momentary assessment of self-reported health-related behaviors and psychological constructs in adults: systematic review and meta-analysis. *J. Med. Internet Res.* 23:e17023. doi: 10.2196/17023

Wrzus, C., & Neubauer, A. B. (2023). Ecological Momentary Assessment: A Meta-Analysis on Designs, Samples, and Compliance Across Research Fields. *Assessment*, *30*(3), 825–846. https://doi.org/10.1177/10731911211067538

Yang, L., Yu, M., & Gao, S. (2016). Joint Models for Multiple Longitudinal Processes and Time-to-event Outcome. *Journal of statistical computation and simulation*, *86*(18), 3682–3700. https://doi.org/10.1080/00949655.2016.1181760

Yen, W., & Fitzpatrick, A. R. (2006). Item response theory. In: R. L. Brennan (Ed.), *Educational Measurement* (4th ed. pp. 111-153). Westport, CN: Praeger Publishers.

Yucel, R. M. (2008). Multiple Imputation Inference for Multivariate Multilevel Continuous Data with Ignorable Non-Response. *Philosophical Transactions of the Royal Society*, 366 (1874): 2389–2403.