THE EFFECTS OF MISSING TIME-VARYING COVARIATES IN MULTILEVEL MODELS

Sierra A. Bainter

A thesis submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Master of Arts in the Department of Psychology.

Chapel Hill
2013

Approved by:

Patrick J. Curran

Daniel J. Bauer

Abigail T. Panter

ABSTRACT

SIERRA A. BAINTER: The Effects of Missing Time-Varying Covariates in Multilevel Models
(Under the direction of Patrick Curran)

Multilevel models are commonly used in psychological research to examine developmentally-motivated hypotheses concerning the within- and between- person effects of a time-varying covariate on some outcome. Whereas multilevel models are flexible to accommodate incomplete data on the outcome, missing time-varying covariates present significant challenges to researchers. Unless multiple imputation is used, missing time-varying covariates will lead to a loss of data. This project evaluated the effects of missing time-varying covariates and imputation of missing time-varying covariates in multilevel models using a multifaceted simulation study. My results showed that missing time-varying covariates can lead to biased parameter estimates. However this bias is likely minor compared to bias already present in complete data due to unreliable estimates of the person-mean of the time-varying covariate. The results presented here are clear motivation for researchers to choose alternative estimation strategies that can account for measurement error in the person mean, whether or not time-varying covariates are missing.

TABLE OF CONTENTS

LIST OF TABLES

Table

LIST OF FIGURES

Figure

CHAPTER 1

INTRODUCTION

Multilevel modeling (MLM) is a widespread technique for analyzing nested data. This flexible method can be used to study individuals nested within groups or repeated measures nested within individual. By modeling repeated measures nested within individuals, researchers can study important issues of inter- and intra- individual change over time (Raudenbush & Bryk, 2002). Time-invariant covariates (TICs) such as gender and race can be incorporated to predict differences in these growth trajectories; in this way it becomes possible to address questions such as "*Who* developed an eating disorder?" and "*Which* adolescents were at risk for drug use?". Similarly, time-varying covariates (TVCs) can be incorporated to explore bivariate relationships between variables over time. TVCs predict time specific deviations from the underlying trajectory, answering questions such as "*When* is disordered eating likely to occur?" and "*When* are adolescents engaging in drug use?"

MLM fluidly extends to situations where the data are unbalanced or the outcome variable is incomplete for some participants (Raudenbush & Bryk, 2002, p.199-200). Whereas a repeated measures ANOVA model requires all subjects to be measured at all time points, the MLM has no such requirement. Furthermore, MLM allows all participants to have different assessment schedules. The estimates of the MLM are chosen by maximum likelihood (ML), meaning the estimation procedure selects values for the model parameters that have the greatest probability of producing the observed data. Because ML is used to estimate the model parameters, MLM will produce unbiased parameter estimates if the outcome variable is incomplete, assuming the missingness is MAR (Enders, 2010, Chapter 4; Raudenbush & Bryk, 2002, p.199-200; Schafer & Graham, 2002). Another widely used missing data technique, multiple imputation (MI; Rubin, 1976), fills in the missing values with probable values while accounting for increased uncertainty in the parameter estimates. MI and ML perform similarly well when the outcome variable is incomplete, but MI is a more involved process (Allison, 2003; Olinsky, Chen, & Harlow, 2003; Schafer, 2003). Altogether, MLM requires a less rigid data structure than classical repeated measures ANOVA, and it also saves the analyst from the extra steps needed to multiply impute the incomplete data.

However, if predictors are missing ML breaks down. Cases with any missing predictors are excluded from analysis, regardless of whether the associated outcomes were observed (Enders, 2010, p. 276-278). This problem is not unique to MLM and also occurs in single-level regression, but this issue is complicated in MLM by the presence of multiple levels of predictors. Specifically, in longitudinal data there are a number of realistic instances where the TVC and outcome could have different patterns of missingness, and the outcome could be observed but not the TVC. For example differential missingness is likely if the TVC and outcome are measured from different reporters. Hussong et al. (2008) encountered this situation when studying the time-varying effects of parent-reported alcoholism on child-reported internalizing symptoms; the parent-reported covariate was missing in cases where the child-reported outcome was observed. In order to avoid deleting observations, they used MI to fill in the missing values before analysis. Similarly this problem could arise if one measure is self-report and another is biological, such as a daily diary study accompanied with a biological monitor of heart rate, hormone levels, or calories burned. TVCs share a dynamic relationship with the outcome, predicting changes both in overall trajectory and in time specific deviations from this trajectory, and it is unclear what specific consequences missing TVCs could have on the estimation of these different effects in the model as well as the variance and covariance components.

Although this limitation of ML is known, the implications of listwise deletion of TVCs have received little attention. A thorough literature review and discussions with experts in missing data and multilevel modeling has revealed that the assumptions made on missing predictors in MLM and the potential consequences of this type of missingness have not been comprehensively studied or explained (however see Shin & Raudenbush 2007, 2010, 2011; and Liu, 2000 for recent exceptions).

The only way to combat listwise deletion in the current MLM framework is to use MI to impute the missing values before analyzing the data. MI is a useful and powerful method, but there are difficulties associated with MI, especially with longitudinal data, which I will describe later. Also, it has yet to be shown that MI will always lead to correct estimates no matter the population-level mechanism for TVC missingness.

In this research project I lay out the assumptions of missing predictors in multilevel models and empirically evaluate the consequences of list-wise deletion as a result of missing TVCs. For the purposes of this project I focus on the issue of missing TVCs, because this is unique to longitudinal data. I study a variety of population mechanisms for TVC missingness, assess their relative impacts on MLM estimates, and determine the extent to which MI will offset bias. This work uniquely assesses the impact of missing TVCs in MLM and is an important exploration of the use of MI in MLM. I also broadened the scope of this work to include useful model extensions that are useful to address developmentally-motivated research questions.

Traditional Multilevel Model

The traditional MLM with random slopes and intercepts is expressed as a set of level-1 and level-2 equations. The level-1 equation is given as

$$y_{ti} = \beta_{0i} + \beta_{1i} time_{ti} + r_{ti} \tag{1}$$

where y_{ti} is the outcome at time *t* for person i, β_{0i} and β_{1i} are the person-specific intercept and slope respectively, and r_{ti} is the person and time specific residual. Time is commonly entered as wave of assessment or chronological age and scaled to the first measurement point so that measured waves 1-5 become $time_{ti}$ = 0-4. The values of time can be set as the chronological age of each participant at assessment, grade in school, days of treatment, or any other time metric useful for the particular analysis question. Nonlinear growth can be modeled by introducing polynomial terms (i.e. include $time_{ti}^2$).

The level-2 equations are

$$\begin{aligned} \beta_{0i} &= \gamma_{00} + u_{0i} \\ \beta_{1i} &= \gamma_{10} + u_{1i} \end{aligned} \tag{2}$$

where γ_{00} and γ_{10} represent the population-average intercept and slope and $u_{0\mathrm{i}}$ and u_{1i} are the person specific deviations of the intercept and slope trajectory. The random slope and intercept growth model equation can be written in reduced-form as

$$y_{ti} = (\gamma_{00} + \gamma_{10} time_{ti}\ ) + (u_{0i} + u_{1i} time_{ti} + r_{ti}). \tag{3}$$

At this point it is helpful to distinguish between fixed and random effects. Fixed effects, denoted with the Greek symbol gamma (γ) are constant across individuals in the population and are the expected, overall average, i.e. $E(\beta_{0i}) = \gamma_{00}$ and $E(\beta_{1i}) = \gamma_{10}$. Random effects are denoted u and vary across individuals. Each person specific u is not actually estimated; rather the model estimates the (co)variance components of the random effects, so that $VAR(u_{0i}) = \tau_{00}$, $VAR(u_{1i}) = \tau_{11}$, and $COV(u_{0i}, u_{1i}) = \tau_{10}$. The variance of the person and time specific residuals can be restricted to be constant over time $VAR(r_{ti}) = \sigma^2$ or allowed to vary by time so that $VAR(r_{ti}) = \sigma_t^2$.

One useful implication of the MLM structure is that the variance of y is partitioned into between and within-person variance. The intra-class correlation (ICC) is the proportion of between-person variance, calculated from an unconditional model as:

$$ICC = \frac{\tau_{00}}{\tau_{00} + \sigma^2} \tag{4}$$

where τ_{00} is the element of **T** corresponding to the variance of u_{0i}. The statistic ranges from 0 to 1, and a high ICC indicates that repeated measures are highly correlated within individual, whereas a low ICC implies that the repeated measures within individual are weakly related. Later in this paper I will discuss an important implication the ICC has on between-person effect estimation.

TICs, person-level predictors such as gender or alcoholism diagnosis, predict between-person differences in the intercepts and slopes. Below a single TIC, denoted x_i, enters the model at level 2. Notice the i subscript to indicate a person-level effect.

$$\begin{aligned} \beta_{0i} &= \gamma_{00} + \gamma_{01}x_i + u_{0i} \\ \beta_{1i} &= \gamma_{10} + \gamma_{11}x_i + u_{1i} \end{aligned} \tag{5}$$

Including TVCs is slightly more complicated; although technically time itself is a time-varying covariate because its value changes from one observation to the next. The effect of an additional TVC can be examined by adding it to the level-1 equation.

$$y_{ti} = \beta_{0i} + \beta_{1i}time_{ti} + \beta_{2i}\, z_{ti} + r_{ti} \tag{6}$$

The corresponding level-2 equations become:

$$\begin{aligned}\beta_{0i} &= \gamma_{00} + u_{0i} \\ \beta_{1i} &= \gamma_{10} + u_{1i} \\ \beta_{2i} &= \gamma_{20} + u_{2i}\end{aligned} \tag{7}$$

The random effect u_{2i}, allows the magnitude of the relation between the TVC and the outcome to vary across individuals. This random effect (or any other random effect) is optionally included. Note that TICs can be included in each of these level-2 equations to explain variability in the intercepts, slopes, or effect of the TVC.

The information captured by a TVC can describe both between- and within-person effects, and simply including the raw TVC at level 1 captures an un-interpretable aggregate effect (Kreft, de Leeuw, & Aiken, 1995; Raudenbush and Bryk, 2002, p. 183). This aggregate effect is weighted composition of the between- and within-person effects. Between-person effects are captured by the mean level of the TVC and within-person effects are the corresponding deviations from this mean (Kreft, de Leeuw, & Aiken, 1995). There are two approaches to disaggregating these effects, and they differ by their approach to centering the TVC, which can either be centered with respect to the grand-mean or with respect to the person-mean (Kreft, de Leeuw, & Aiken, 1995; Raudenbush and Bryk, 2002, p. 183). The simplest method for interpretation is to person-mean center by including the person specific mean at level 2, denoted $\bar{z}_i$, and including the person-mean centered TVC, $\dot{z}_{ti}$ at level 1. The set of equations becomes:

Level 1:

$$y_{ti} = \beta_{0i} + \beta_{1i} time_{ti} + \beta_{2i}\dot{z}_{ti} + r_{ti} \tag{8}$$

Level 2:

$$\begin{aligned}\beta_{0i} &= \gamma_{00} + \gamma_{01}\bar{z}_i + u_{0i} \\ \beta_{1i} &= \gamma_{10} + u_{1i} \\ \beta_{2i} &= \gamma_{20} + u_{2i}\end{aligned} \tag{9}$$

Reduced-form:

$$y_{ti} = (\gamma_{00} + \gamma_{01}\bar{z}_i + \gamma_{10}time_{ti} + \gamma_{20}\dot{z}_{ti}) + (u_{0i} + u_{1i}time_{ti} + u_{2i}\dot{z}_{ti} + r_{ti}) \quad (10)$$

As Curran and Bauer (2011) point out, this approach to estimating TVC effects assumes that the TVC is unrelated to the passage of time. If there is growth in the TVC over time, additional adjustments are needed to obtain correct between- and within-person effect estimates.

By person-mean centering in this way the between- and within-person effects of the TVC are untangled and can be examined separately. This is important because a TVC can exhibit (1) both a between-person effect and a within-person effect (operating in the same or opposite directions), (2) a between-person effect but no within-person effect, (3) a within-person effect but no between-person effect, or (4) neither type of effect (see Curran & Bauer, 2011 for a discussion on disaggregating effects).

The model can be extended even further to allow one or more TICs to predict differences in the effect of the TVC or by including interactions. For example, the effect of the TVC could differ for men and women, or it could differ depending on the overall level of the TVC, $\bar{z}_i$. The effect of the TVC can even interact with time, meaning the magnitude of the relation between the TVC and the outcome changes with time. These extensions are included below to create a final, general model.

Level 1:

$$y_{ti} = \beta_{0i} + \beta_{1i}time_{ti} + \beta_{2i}\dot{z}_{ti} + \beta_{3i}\dot{z}_{ti}time_{ti} + r_{ti} \quad (11)$$

Level 2:

$$\begin{aligned} \beta_{0i} &= \gamma_{00} + \gamma_{01}\bar{z}_i + \gamma_{02}x_i + u_{0i} \\ \beta_{1i} &= \gamma_{10} + \gamma_{11}\bar{z}_i + \gamma_{12}x_i + u_{1i} \end{aligned} \quad (12)$$

$$\beta_{2i} = \gamma_{20} + \gamma_{21}\bar{z}_i + \gamma_{22}x_i + u_{2i}$$
$$\beta_{3i} = \gamma_{30} + \gamma_{31}\bar{z}_i + \gamma_{32}x_i + u_{3i}$$

Reduced-form:

$$\begin{aligned} y_{ti} = (&\gamma_{00} + \gamma_{01}\bar{z}_i + \gamma_{02}x_i + \gamma_{10}\, time_{ti} + \gamma_{11}\bar{z}_i\, time_{ti} + \\ &\gamma_{12}x_i time_{ti} + \gamma_{20}\dot{z}_{ti} + \gamma_{21}\bar{z}_i\dot{z}_{ti} + \gamma_{22}x_i\dot{z}_{ti} + \\ &\gamma_{30}\dot{z}_{ti}time_{ti} + \gamma_{31}\bar{z}_i\dot{z}_{ti}time_{ti} + \gamma_{32}x_i\dot{z}_{ti}time_{ti}) + \\ &(u_{0i} + u_{1i}time_{ti} + u_{2i}\dot{z}_{ti} + u_{3i}\dot{z}_{ti}time_{ti} + r_{ti}) \end{aligned} \tag{13}$$

This expanded model includes fixed effects for the overall-intercept (γ_{00}), between-person effect of the TVC (γ_{01}), between-person effect of the TIC (γ_{02}), overall-slope (γ_{10}), effect of the overall level of the TVC on the slope (γ_{11}), effect of the TIC on the slope (γ_{12}), overall within-person TVC effect (γ_{20}), effect of the overall level of the TVC on the within-person TVC effect (γ_{21}), effect of the TIC on the within-person TVC effect (γ_{22}), overall TVC by time interaction (γ_{30}), effect of the overall level of the TVC on the TVC by time interaction (γ_{31}), and effect of the TIC on the TVC by time interaction (γ_{32}). The random effects for the intercept (u_{0i}), slope (u_{1i}), TVC (u_{2i}), and TVC by time interaction (u_{3i}) allow all of these effects to vary across individuals. Finally, the level-1 residual (r_{ti}) accounts for additional unexplained variability.

As with any statistical model, a number of assumptions underlies all of these models. I will detail these in the next section.

Model Assumptions

The assumptions of the MLM are similar to those for regression with some necessary extensions and with assumptions applied to each level of the model. At level 1, the residuals are assumed to be conditionally independent and identically distributed both within and across individuals. Formally this assumption is written as $r_{ti} \sim$ iid $N(0, \sigma^2)$. The assumption that the level-1 residual variance is equal at all time points is referred to as homoscedasticity and can be relaxed if needed at the cost of some parsimony. In longitudinal data analysis it is often necessary to relax the assumption of homoscedasticity; for example, it is common to observe that the residual variance increases with time.

At level 2, all q random effects are assumed to be conditionally independent across levels and multivariate normally distributed with mean equal to zero and covariance matrix **T**, or: $u_{qi} \sim$ iid $MVN(0, \mathbf{T})$. The random effects and residuals are also assumed uncorrelated across levels of the model, so that $\mathrm{Cov}(u_{qi}, r_{ti}) = 0$ for all random effects q.

All predictors, whether level-1 or level-2, are assumed to be independent of r_{ti} and u_{qi}. This means that $\mathrm{Cov}(z_{ti}, u_{qi}) = 0$ and $\mathrm{Cov}(z_{ti}, r_{ti}) = 0$ for all q random effects and likewise $\mathrm{Cov}(x_{si}, u_{qi}) = 0$ and $\mathrm{Cov}(x_{si}, r_{ti}) = 0$ for all s predictors.

Furthermore, predictors are assumed fixed and known and thus without measurement error. This means that no assumption is made about the distribution of the predictors. Instead, distributions are assumed for r_{ti} and u_{qi}. Taken together, the above model structure implies a marginal distribution for y and values for the means and (co)variance components of the observations within and between individuals, and all of these distributions are conditional on the fixed values of the predictors. The assumption that the predictors are treated as fixed and known in particular is a key focus in this project.

Bias in the Between-Person Effect

Recently the issue sampling error of the group-mean, or in this case person-mean, has been gaining attention in the multilevel literature. This was most clearly addressed by Lüdtke et al. (2008) where bias due to this sampling error is demonstrated mathematically and with comprehensive simulation results (see also Curran et al., in press; Hoffman & Stawski, 2009;

Marsh et al., 2007; Preacher et al., 2011; and Shin & Raudenbush, 2010). The problem is the traditional multilevel model does not account for measurement error in the predictors; instead predictors are treated as fixed and known. Estimating the person mean is necessary to disaggregate between- and within-person effects, but there is error in this person mean just as error is expected in a sample mean. In other words, the person-specific mean is used as a predictor (fixed and known), but the person-specific variability around this mean is ignored. Ignoring this variability around the person-mean assumes that the mean is measured without error for all persons.

Sampling error in the person mean leads directly to bias in the between-person effect. The extent of this bias depends on the number of repeated measures, the ICC of the TVC, and the magnitude and direction of the between- and within-person effects. Lüdtke et al. (2008) gave a formula for the bias, assuming an equal number of repeated measures for all individuals. Below we see that the expected difference between the estimated between-person effect ($\hat{\gamma}_{01}$) and the true effect ($\beta_{between}$) will be in the direction of the true within-person effect (β_{within}):

$$E(\hat{\gamma}_{01} - \beta_{between}) = (\beta_{within} - \beta_{between}) \cdot \frac{1}{T} \cdot \frac{(1 - ICC)}{ICC + (1 - ICC)/T} \qquad (14)$$

This formula also shows that a crucial element of this bias is the number of repeated measures, *T*. When the number of repeated measures is small, as is often the case in longitudinal studies, this bias can potentially be extreme, even in complete data. Simulation studies by Lüdtke et al. (2008) and others (e.g. Marsh et al., 2009) have shown this bias in group-level data, where often the group size is much larger than the number of repeated measures collected in a longitudinal study; however, less attention is paid to the strong potential for bias in repeated measures data (but see Curran et al., in press; Hoffman &

Stawski, 2009). The extent of the bias can only be known in simulated data when the population parameters are known, but the direction of the within-person effect will indicate the direction of the bias.

So far I have only considered the implications of between-person effect bias if the data are completely observed and all individuals have the same number of repeated measures. If some measures of the TVC are missing, this will likely increase sampling error in the person mean. In all cases, if the between-person effect is biased, this bias will spread to related estimates in the model (Lüdtke et al., 2008). As I discussed previously, TVCs have the potential to be unobserved in a study, so the implications of missing TVCs should be better understood. In the next section I provide a brief overview of missing data theory.

Overview of Missing Data Theory

The foundation of missing data theory and the widely accepted taxonomy of missing data mechanisms was originally described by Rubin (1976). The key idea behind Rubin's theory is that missingness is a variable with a probability distribution. Define a binary variable *R* as a missing data indicator that takes on a value of 1 when a value is observed and 0 when missing. In the case of multivariate missingness the missingness indicator *R* becomes a matrix, **R**. The probability distribution of *R* is the probability of missing data, and this probability may or may not be related to other variables in the data. The hypothetical complete data (Y_{com}) is partitioned into observed and unobserved portions (Y_{obs} and Y_{miss}), and we want to understand the $p(R|Y_{obs}, Y_{miss}, \theta)$, where θ is a parameter or set of parameters that defines the relationship between *R* and the data. Note that *Y* in this notation represents all of the data, not just the outcome variable. The three missing data mechanisms are missing at random (MAR), missing completely at random (MCAR), and missing not at random (MNAR). I will briefly explain each in turn.

If the data are MAR then the probability of missingness on a variable Y is related to another variable (or variables) in the analysis but is unrelated to the missing values of Y. For example, if the probability of missingness is higher for minority groups, the assumption of MAR will be met as long as minority status is included in the analysis. Similarly, missingness could relate to time such that individuals are less likely to be observed at later time points. In the case of MAR missingness, the probability of missingness reduces to:

$$p(R|Y_{obs}, Y_{miss}, \theta) = p(R|Y_{obs}, \theta) \tag{15}$$

MCAR is a special case of MAR, and is the mildest form of missingness. If data is MCAR, this means the probability of missingness is unrelated to any values in the data, either observed or unobserved. The probability expression reduces to:

$$p(R|Y_{obs}, Y_{miss}, \theta) = p(R|\theta) \tag{16}$$

This assumption is the most stringent, and MCAR data are rare in practice. Important exceptions are planned missingness designs and completely arbitrary losses of data.

Finally, if data are MNAR, this means the probability of missingness depends on the missing values themselves, and the probability expression cannot be reduced.

$$p(R|Y_{obs}, Y_{miss}, \theta) = p(R|Y_{obs}, Y_{miss}, \theta) \tag{17}$$

Importantly, it is almost always impossible to know the mechanism for missingness in real data (Enders, 2010, pp. 17-21). Different procedures for dealing with missing data are able to handle different types of missingness. I will briefly describe listwise deletion, an outdated technique that is unfortunately still in use and then move to multiple imputation (MI) and maximum likelihood (ML) which are considered to be the current “state of the art” when dealing with data assumed to be at least MAR (Schafer & Graham, 2002). To limit my focus, I will exclude techniques for handling MNAR data.

Listwise Deletion

Until recently researchers had little choice but to use a variety of outdated techniques for handling missing data, none of which are generally advisable. It is not my intention to comprehensively list or argue against these techniques here, but see Enders (2010) and Schafer and Graham (2002) for more details.

Of interest to this project is the practice of listwise deletion, also called complete-case analysis. Listwise deletion is often the easiest way to handle missing data by simply excluding it from analysis. This reduces sample size and power and requires the restrictive missingness assumption of MCAR. Use of this outdated technique is unwarranted, especially given that more appropriate alternatives are available. Superior techniques like ML and MI do not require the unrealistic and unnecessarily strict assumption of MCAR and can avoid wastefully discarding data.

Maximum Likelihood

There is no closed-form mathematical formula to estimate all MLM parameters in realistic applications. Maximum likelihood estimation (see Rubin, 1987, section 6.2 for computational details) is therefore generally used via an iterative algorithm to estimate the model, and common software automatically implements ML or Restricted Maximum Likelihood (REML) estimation. When generalized to the case of incomplete data, ML estimation is asymptotically unbiased and efficient, provided the missingness is MAR and the model is reasonable for the data. ML is therefore considered one of two “state-of-the-art” (Schafer & Graham, 2002) approaches for handling missing data. Because ML makes use of all the available data it will outperform listwise deletion even if data are MCAR (Enders, 2001).

The point of maximum likelihood estimation is to choose the population parameters that maximize the likelihood of observing a particular sample (for a straightforward overview of ML, see Enders, 2010, Chapters 3 & 4). This process begins by specifying a distribution for the population, and in the social sciences this is usually univariate or multivariate normal. In the multivariate case the distribution is expressed by a probability density function which describes the height of the curve L_i for the vector of scores $\mathbf{Y}_i$ obtained for case i. For each $\mathbf{Y}_i$ the value of L_i gives the relative probability of observing a particular set of scores from a normally distributed population with the specified population mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$.

Assuming independent observations, the likelihood of the sample is the product of the likelihoods for each score, and the goal of maximum likelihood estimation is to maximize this likelihood by auditioning different values for $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$. In practice it is easier to maximize the log-likelihood to avoid rounding error and because the sample log-likelihood is the sum (rather than the product) of the individual log likelihood values.

Assuming a multivariate normal distribution for the data, the equation for the complete data individual log likelihood is:

$$\mathrm{logL_i} = -\frac{k}{2}\log(2\pi) - \frac{1}{2}\log|\boldsymbol{\Sigma}| - \frac{1}{2}(\mathbf{Y_i} - \boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\mathbf{Y_i} - \boldsymbol{\mu}) \qquad (18)$$

The term $(\mathbf{Y_i} - \boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\mathbf{Y_i} - \boldsymbol{\mu})$ gives the standardized distance between each score and the mean. Called the Mahalanobis distance, this term determines the relative magnitude of each log likelihood value. To the left of the Mahalanobis distance is a collection of scaling terms that make the area under the distribution sum to one.

ML estimation easily generalizes to accommodate incomplete data. The equation for the individual multivariate log likelihood for incomplete data is:

$$\log \mathrm{L_i} = -\frac{k_i}{2}\log(2\pi) - \frac{1}{2}\log|\boldsymbol{\Sigma}_\mathrm{i}| - \frac{1}{2}(\mathbf{Y_i} - \boldsymbol{\mu}_\mathrm{i})^\mathrm{T}\boldsymbol{\Sigma}_\mathrm{i}{}^{-1}(\mathbf{Y_i} - \boldsymbol{\mu}_\mathrm{i}) \qquad (19)$$

The only difference between the two equations is subscripting of the parameter matrices. These matrices are now allowed to conform to the dimensions needed to describe the complete data for each case.

Software programs choose starting values for the estimates and then use iterative optimization algorithms to adjust the estimates until "convergence" is reached. Convergence occurs when the change in log-likelihood from iteration to iteration becomes essentially zero; at this point the log-likelihood is maximized and the algorithm has arrived at the maximum likelihood estimates. ML estimation extended to incomplete data is virtually automatic to estimating the parameters in MLM software, and is therefore understandably the missing data procedure of choice for most researchers conducting MLM. Unfortunately, ML estimation is not without its limitations.

Limitations of Maximum Likelihood in Multilevel Models

There is an important instance when ML is inadequate. ML estimation is well-suited for situations where the outcome is missing, but when predictors are missing this estimation is not possible. To see why ML cannot proceed with missing predictors, look again at Equation (19) for the individual log likelihood when missing data is present. Implied by this equation is that a distribution is specified for the outcome variable, but nowhere is a distribution specified for predictors. When predictors are present, the $\boldsymbol{\mu}_\mathrm{i}$ in this equation refers to the conditional mean, or predicted score, of the outcome. It is simple to infer values for the outcome once we have (a) made an assumption about its distribution and (b) know the values of the predictors. It is useful to note that this is the meaning of the MLM assumption stated previously that the predictors are *assumed fixed and known*.

The above explanation is simple, but the severe consequence is any cases with missing predictors in MLM are excluded from analysis (Enders, 2010, p. 276-278). Even if Y is present, it will be excluded as well, and this is analogous to the outdated practice of listwise deletion of observations. This problem is not isolated to MLM; it is a fundamental feature of conditional likelihood. What *is* a unique issue to MLM is the presence of multiple levels of predictors. If a TIC (person-level covariate) is missing, that entire subject is excluded from analysis. Missing TVCs will exclude the associated wave for each subject from analysis, and this introduces some interesting complexities.

Not only does this omit a wave of data, it also decreases the amount of data available to compute the person-mean of the TVC and systematically missing TVCs will systematically bias the person mean estimate. Since the person-mean is prone to measurement error (Lüdtke et al., 2008), even unsystematic TVC missingness (that is TVCs that are MCAR) will increase measurement errors, and this is a non-trivial consequence which has yet to be studied. Further, in the case of systematically missing TVCs, missing due to an MAR or MNAR mechanism, will systematically bias the person mean. Many factors besides mechanism (i.e. MAR versus MNAR) will systematically impact the mean in different ways. In other words, the exact form of the missingness pattern will dictate how the mean is systematically biased and there are countless possible permutations. For example, in the simple case that higher values of TVCs are more likely to be missing, person mean estimates will be downwardly biased. Listwise deletion of cases with missing TVCs in order to use ML estimation will not counteract this bias. Despite the convenience of ML for MLM, it is not an appropriate method for handling missing data if predictors are incompletely observed.

Multiple Imputation

The other widely used state-of-the-art technique for handling missing data, MI was proposed by Rubin (1987). Like ML, MI assumes MAR missingness and multivariate normality. Whereas ML estimates the parameters directly from the available data, MI creates complete data sets by filling in the missing values with plausible values. This process is repeated over *m* replications, creating *m* complete data sets. Each data set is analyzed separately, and the estimates and standard errors are combined using simple rules given by Rubin (1987).

The imputation phase of multiple imputation cycles between two steps. In the first step, the incomplete data are predicted from the complete data using regression equations based on the estimated means and covariances. Each predicted score is augmented with a residual drawn from the appropriate normal distribution with variances equal to the residual variance from the regression. The second step depends on Bayesian methodology to generate different estimates for the regression coefficients and ultimately to arrive at *m* plausible alternative versions of the complete data.

Distinct from other Monte Carlo procedures, traditionally only a small number of imputations, 5 or 10, have been recommended (Rubin, 1987; Schafer, 1997). Recently methodologists have begun to recommend more imputations to improve power (see Graham, Olchowski, & Gilreath, 2007). The recommended number of imputations varies depending on the rate of missing information, but 20 is the current rule of thumb for most situations.

MI of multilevel data is more complicated than with single-level data. If single-level imputation methods are applied to a multilevel dataset, the method assumes the error variance for a variable is homogeneous across time points (Schafer, 2001). Multilevel imputation methods rely on a particular iterative algorithm called the Gibbs sampler (see Casella & George, 1992), and currently these methods are not available in commercial statistical packages. In some cases, an easy fix is to impute the data in wide format (one record per subject). If the dataset is in wide format, single-level imputation allows the error variance to be heterogeneous across time points (Fitzmaurice, Laird, & Ware, 2011, p. 543). This is only appropriate if the data are rigidly structured, meaning all subjects follow the same assessment schedule. If this is not the case, imputing the data in wide format induces additional missingness and single-level imputation procedures will not work properly. For this masters project I simulated data, so the complications of unstructured data did not arise.

Multiple Imputation Versus Maximum Likelihood

It is well established that when the same variables are used in the model and when sample size is large, ML and MI yield practically identical results (e.g., see Allison, 2003; Newman, 2003; Olinsky, Chen, & Harlow, 2003; Schafer, 2003). However, both methods handle missingness in distinctively different ways, and they each have their advantages. In many instances, ML is easier to implement because it is not necessary to go through all of the steps required in MI before doing an analysis. However there are instances where MI should be preferred. For example, as Enders and Gottschall (2011) point out, if the ultimate goal of an analysis is to compute a scale score for subsequent analysis steps, only multiple imputation supplies the filled in item-level scores necessary to do this. Similarly, in the case of missing TVCS, MI can preserve the item-level TVCs necessary for computing the person mean.

Additionally, MI makes it is easier to include auxiliary variables, which are potential correlates of missingness or the incomplete variables that are not of direct theoretical interest (Collins, Schafer, & Kam, 2001). Including these auxiliary variables can be helpful for satisfying the MAR assumption. Procedures outlined to incorporate auxiliary variables into ML analyses are currently cumbersome and infrequently used (see Enders, 2010). This is not an advantage of MI per se but is more a feature of the statistical software currently in use. It is reasonable to expect software will catch up and counteract this imbalance.

Finally, in the case of missing predictors in MLM, MI clearly holds the advantage over ML. MI is equally useful for missing predictors and missing outcomes.

Other Alternatives

The approaches I have presented here represent the traditional multilevel model and the options available within this framework. Other less widely available alternatives exist. Recently, Shin and Raudenbush (2007, 2010, & 2011) have proposed alternative MLM approaches that allow the model to be estimated using the joint likelihood instead of the conditional likelihood as described previously. Using this approach, listwise deletion of TVCs is avoided. This development is quite recent and not currently programmed in any commercially available software.

Joint likelihood estimation is also potentially available for analyzing longitudinal data in the structural equation modeling (SEM) framework. However, when interested in separating between- and within-person effects, there is no straightforward analogue between a traditional MLM growth curve and an SEM latent growth curve (Curran et al., in press). In this project I limited my focus to answer this question in the traditional MLM context because (1) this approach is in wide use and (2) the issue remains that the impact of missing TVCs has not been appropriately studied.

Research Hypotheses

Based on prior work by Lüdtke et al. (2008) and pilot simulation work I expected that TVCs missing by any mechanism would increase sampling error in the person-mean, and this would cause bias that would propagate throughout the model estimates. This bias due to sampling error in the person-mean should impact estimates of the fixed effects for the intercept, slope, between-person effect, and possibly interaction effects. The elements of **T** should be sensitive to this error as well. With more repeated measures, the effects of TVC missingness should be less extreme. When substantial bias is present, it is possible that spurious effects may be detected. The within-person effect of the TVC and person-level variance σ^2 were expected to remain stable. The effects of sampling error in the person-mean on more complex interactions with $\bar{z}_i$, $\dot{z}_{ti}$ and TICs had not been previously examined, and I did not have prior hypotheses about these effects. Some evidence from pilot simulations suggested that a spurious $time_{ti}$*$\dot{z}_{ti}$ interaction may result from increased sampling error in the person-mean.

Multiple imputation of missing TVCs is predicted to decrease bias, though not when TVCs are MNAR, and I do not expect MI will improve bias beyond the levels exhibited in complete data. When TVCs are missing by any mechanism besides MCAR, this is expected to systematically bias the person mean. The direction and extent of this bias will depend not only on whether the missingness is MAR or MNAR, but on precisely how the mechanism will affect the person mean via which types of TVC observations are more likely to be missing.

CHAPTER 2

METHOD

To evaluate the impact of missing TVCs in multilevel models I carried out a computer simulation study. I simulated data consistent with a longitudinal multilevel model, systematically imposed missingness, and multiply imputed the missing data. In this design I varied the percentage missing, the missingness mechanism, and the number of repeated measures. All data generation and analyses were done in SAS.

Data Generation and Model Design

The simulated data were consistent with a longitudinal growth model with a TVC (e.g., Equation 10). The ICCs of the TVC and the outcome were set to .50 when computed from unconditional models, a sufficiently high ICC to be reflective of longitudinal data. For simplicity, the level-1 error structure was homoscedastic, but this does not limit the generalizability of the findings.

Many simulation studies focus on an artificially simple model, but in the present study I constructed more realistically complex models. As I have shown before, a variety of exciting hypotheses can be explored with TVCs in multilevel models, and I chose to evaluate the impact of TVC missingness on these motivating examples in order to have more generalizable results. To do this I examined a basic TVC model (Model 1) and then built up to a more interesting population generating model (Model 2). To ground the models to a substantive framework, the motivating example was the hypothesized effects of internalizing symptoms on alcohol use for male and female adolescents.

Model 1. Model 1 included random intercept and random slope (time) effects, and the TVC (e.g., internalizing) exerted both within- and between-person effects on the outcome (e.g., alcohol use). A binary TIC (e.g., gender) also asserted an intercept effect (i.e., a main effect on the outcome).

The level-1 and level-2 equations for Model 1 are shown below.

Model 1:

Level 1:

$$y_{ti} = \beta_{0i} + \beta_{1i}time_{ti} + \beta_{2i}\dot{z}_{ti} + r_{ti} \tag{20}$$

Level 2:

$$\begin{aligned} \beta_{0i} &= \gamma_{00} + \gamma_{01}\bar{z}_i + \gamma_{02}x_i + u_{0i} \\ \beta_{1i} &= \gamma_{10} + \gamma_{11}\bar{z}_i + u_{1i} \\ \beta_{2i} &= \gamma_{20} \end{aligned} \tag{21}$$

These population values were chosen to be representative of effects found in real applications. The intercept (γ_{00}) was set to 5.00 with a TIC intercept effect $\gamma_{02} = 1.00$ which represents that females started one unit (or 20%) lower than males. The between- (γ_{01}) and within- (γ_{20}) person effects representing the effects of internalizing on alcohol use are in the same direction, where $\gamma_{20} = 1.00$ and $\gamma_{01} = 0.20$. This general pattern of within- and between-person effects (in the same direction with a larger within-person effect) is a pattern often seen in real data (e.g. Armeli et al., 2005; Hardy et al., 2011; Jahng et al., 2011). The slope of alcohol use increased at a moderate rate of one half unit per wave, $\gamma_{10} = 0.50$. Note that in these equations a term for the overall level of the TVC interacting with time (γ_{11}) is included, but the population value for this parameter is zero.

The elements of **T** and the residual variance were all chosen to reflect credible variability. Consistent with MLM assumptions, the residual variances were generated r_{ti}~ iid $N(0, \sigma^2)$ where σ^2 = 2.0. Likewise the random effects were multivariate normally distributed, or: u_{qi}~ iid $MVN(0, \mathbf{T})$. Within **T**, the intercept variance (τ_{00}) was 2.00, slope variance (τ_{11}) was 1.00, and the intercept and slope covariance (τ_{01}) was -0.20 (with associated correlation -0.10), reflecting the common pattern that, in general, those who start higher increase at a slightly slower rate.

Model 2. Model 2 included all of the effects in Model 1. Additionally, Model 2 included a random within-person effect for the TVC and an interaction between the within-person TVC effect and time. Applications with random within-person TVC effects are not typically tested, but interesting examples have been published. In one recent example, Stawski, Mogle, & Sliwinski (2011) included a random within-person daily stressor effect to test for individual differences in this effect. The level-1 and level-2 equations are shown below in Equations 22 and 23.

Model 2:

Level 1:

$$y_{ti} = \beta_{0i} + \beta_{1i} time_{ti} + \beta_{2i}\dot{z}_{ti} + \beta_{3i}\dot{z}_{ti} time_{ti} + r_{ti} \tag{22}$$

Level 2:

$$\begin{aligned} \beta_{0i} &= \gamma_{00} + \gamma_{01}\bar{z}_i + \gamma_{02}x_i + u_{0i} \\ \beta_{1i} &= \gamma_{10} + \gamma_{11}\bar{z}_i + u_{1i} \\ \beta_{2i} &= \gamma_{20} + u_{2i} \\ \beta_{3i} &= \gamma_{30} \end{aligned} \tag{23}$$

The interaction term γ_{30}= 0.05 included in Model 2 signifies that the magnitude of the within-person TVC effect increased modestly with time; this effect is included because it is relevant to developmental processes. The variance for the random within-person effect (τ_{22}) was 0.10, and this random effect did not covary with the intercept or slope (i.e., $\tau_{20} = \tau_{21} = 0$).

Sample Size and Number of Repeated Measures

Because sample size is unrelated to between-person effect bias and was not expected to play an important role in the conditions I examined, I held sample size constant at N=250 per data set. This relatively large sample size is representative of a large-scale study and was also chosen to help avoid any potential convergence issues with MI. In pilot work, MI had more difficulty converging in smaller sample size conditions. However, these convergence issues are nontrivial, and with real data it is impossible to simply increase sample size as can be done with artificial data. The purpose of my thesis was not to investigate convergence issues in MI, but more research in this area is needed to comprehensively ensure that MI for TVC missingness is viable in smaller samples.

The number of repeated measures was varied at 5 and 10. This is an important factor to vary because the number of repeated measures directly impacts the amount of sampling error in the person mean and should impact the magnitude of bias caused by list-wise deletion of TVCs. Both of these conditions are representative of traditional multi-wave designs.

Proportion of Missingness

I examined two rates of TVC missingness, 20% and 40%; all other variables were completely observed. The 20% missing condition was intended to represent a situation that could likely occur in practice. The 40% missingness condition is more extreme, but still not unreasonable, and this condition is worth examining to see what may happen in a severe case. All data were completely observed at wave 1, a realistic assumption for longitudinal data. Therefore these rates of missingness apply to all subsequent waves. For example, with 5 repeated measures and 20% missing MCAR, wave 1 was completely observed and waves 2-4 had 20% missing from each wave.

Missingness Mechanisms

Each condition started with the same complete simulated data sets. TVCs from each complete data set were deleted consistent with an MCAR, MAR, and MNAR missingness mechanism in turn. The forms of missingness imposed here are intermittent, meaning a participant can have any pattern of missing and observed waves after the first time point. This is in contrast to the less general case of dropout missingness, where a participant never re-enters the study after the first missing wave.

To impose MCAR missingness, I simulated a random normal value for each TVC and ranked this random variable into 10 groups by wave. I then deleted TVCs from each wave (starting at wave 2) that corresponded to the appropriate number of highest groups for each percent missing condition (i.e., deleted 4 highest groups for 40% missingness condition). MNAR and MAR missingness was related to the values of the TVCs themselves such that higher values of the TVC were deleted, based on either the current TVC value (MNAR) or previous TVC value (MAR). Specifically, for MNAR missingness, the highest 20 or 40% of TVC values were deleted, by sample and wave. For MAR missingness, TVCs were deleted based on the highest 20-40% of previously observed values.

Analysis

After deleting TVCs, each incomplete data set was analyzed first using PROC MIXED (i.e. listwise deletion) in SAS and again after MI. Because I was not interested in examining model bias due to omitted effects or model misspecification, all models were correctly specified to correspond to either Model 1 or 2. However, bias in the model can lead to the detection of spurious effects, so to detect this I included a time by $\bar{z}_i$ interaction in the model statement even though this interaction did not exist in either population generating model. In practice non-positive definite solutions are not retained, therefore to maximize external validity, replications with non-positive definite solutions were discarded.

After evaluating the impact of TVC missingness, each simulated data set was multiply imputed using SAS PROC MI. In order to preserve the nested mean and correlation structure, each data set was transposed from long format to wide format before imputation. As recommended for intermittent missingness, the MCMC method was used with 1000 burn-in iterations for each chain (Schaeffer, 1997). The initial mean and covariance estimates were derived for a posterior mode from the EM algorithm, which was also allowed a maximum of 1000 iterations. Following the recommendations of Graham et al. (2007), I carried out 20 imputations per data set. In certain conditions with high MCAR missingness and repeated measures, the posterior covariance matrices for MCMC were singular, and these replications were discarded. Information about discarded replications is presented in the results section.

After MI, the data were transposed back into long format and analyzed as before using SAS PROC MIXED. To be conservative and to avoid comparing results based on different numbers of imputations, replications in which any of the 20 imputations yielded a non-positive definite solution were discarded, and I will present more information about discarded replications with my results. Finally I used SAS PROC MIANALYZE to combine the estimates and standard errors for each data set.

Evaluation criteria. Within each condition, the performance of the model was evaluated by computing raw bias, relative bias, and root-mean-square error (RMSE). The proportion of significant effects was examined for estimates that were zero in the population;

this is the observed Type I error rate. Raw bias is simply the true value (θ) subtracted from the corresponding estimate ($\hat{\theta}$)

$$Raw\ Bias = \hat{\theta} - \theta. \tag{24}$$

Relative (or percentage) bias is computed as

$$Rel\ Bias = \frac{\hat{\theta} - \theta}{\theta} \times 100 \tag{25}$$

and is simply raw bias scaled as a percentage of the population parameter. For the parameters that equal zero in the population model, such as the interaction between time and the between-person TVC effect, this statistic is not defined because it would involve dividing by zero. RMSE is computed as

$$RMSE = \sqrt{\frac{\sum(\hat{\theta} - \theta)^2}{N}} \tag{26}$$

and because it is in the same metric as the data, it can be interpreted as representative of the size of a "typical" error. RMSE is not strictly a measure of bias; rather it takes into account the variance of the errors and the mean error, so RMSE will not necessarily be zero when a parameter estimate is unbiased.

The factors of the simulation were examined for each parameter using analysis of variance (ANOVA) models with raw bias as the outcome, separately for Model 1 and Model 2. These meta-models were used to identify the most important factors in the simulation for explaining bias in each parameter. First, I analyzed the complete data separately for Model 1 and Model 2 using a one-factor ANOVA to test the effect of five versus 10 repeated measures on relative bias in each parameter estimate. In the missing data conditions for Model 1 and Model 2, four-factor 2x2x2x3 ANOVAs were examined to detect significant interaction and main effects among the factors manipulated: estimation method (Listwise vs

MI), number of repeated measures (5 vs 10), proportion of missingness (20 v 40%), and missingness mechanism (MCAR, MAR, MNAR).

Because ANOVA has high power to detect significant effects, I used partial η^2values of each effect as an effect size measure to screen for meaningfully large effects. Partial $\eta^2$is computed as

$$Partial\,\eta^2 = \frac{ss_{Between}}{ss_{Between} + ss_{Within}} \tag{27}$$

where $ss_{Between}$ and ss_{Within} are the sums of squared deviations from the mean, representing between-group and within-group variability respectively. Corresponding to a conventional medium effect size (Cohen, 1988), I examined significant effects that produced a partial η^2value of at least .06. Finally, I followed up appropriate main effects and interactions using graphics. RMSE, relative bias, and proportion of significant effects were examined as descriptive measures.

CHAPTER 3

RESULTS

First, I will present results from the complete data conditions for each model to validate that the data generation was correct and to examine bias in complete data. In the complete data conditions, I examined partial η^2 values from the ANOVA meta-models to assess the impact of the number of repeated measures on raw bias in the parameter estimates. I show that this bias is consistent with what we would predict from Lüdtke et al. (2008). Descriptively, I also examined RMSE and the proportion of Type I errors for effects that were zero in the population generating models. Second, I will present results from the missing data conditions. Once again, I examined partial η^2 values from ANOVA meta-models to detect meaningful main effects and interactions among the simulation factors. To untangle the interaction effects, I examined relative bias because of its meaningful metric as well as RMSE.

Complete Data Conditions

The ANOVA meta-model results for complete data are shown in Table 1 along with the recovery of the population generating values for the complete data conditions. In the complete data, the only comparison was the effect of five versus 10 repeated measures on relative bias. In Model 2, seven solutions were discarded in the five repeated measures conditions for non-positive definite estimates of the random within-person effect. All other complete data replications converged to proper solutions.

Model 1. As expected, there was considerable bias in the complete data, especially for the intercept (γ_{00}) and between-person TVC effect (γ_{01}), as shown in Table 1. In the ANOVA meta-model, there were substantial main effects for the repeated measures factor on the intercept and between-person TVC effect. With more repeated measures, bias decreased. The intercept was underestimated by about 13% and 11% relative bias for 5 and 10 repeated measures, corresponding to sizable RMSE values of .66 and .55.

In terms of raw bias, the between-person TVC effect estimate was overestimated by .13 with 5 repeated measures and .07 with 10 repeated measures, exactly as predicted by Equation 14 given the population generating values. Bias in the between-person effect corresponded to large relative bias, 66% and 35% relative bias for the 5 and 10 repeated measures conditions, respectively. RMSE, reflecting variability and bias, for the between-person TVC effect was .13 for 5 repeated measures and .09 for 10 repeated measures.

Model 2. The pattern of complete data results were essentially the same in Model 2 as observed in Model 1. The meta-model confirmed sizable effects of the number of repeated measures on intercept and between-person effect estimates. In terms of relative bias, the intercept was underestimated by about 13% with five repeated measures, or 0.66 RMSE. With 10 repeated measures, the bias decreased to 11%, or .55 RMSE. Similarly, relative bias in the between-person effect decreased from .33 with 5 repeated measures to .27 with 10 repeated measures.

Missing Data Conditions

I will present results for the missing data conditions separately for each model, although, as I will show, many findings are similar across models. Because of the high power to detect effects in the meta-models for each parameter, I only examined main effects and interactions with partial eta-squared values of .06 or larger, corresponding to a medium effect size.

Model 1. Meta-model results for Model 1 are summarized in Table 2. Meta-model results indicated that there were nontrivial effects for the intercept (γ_{00}), between-person TVC effect (γ_{01}), within-person TVC effect (γ_{20}), and residual variance (σ^2). There were no effects of any factor on intercept variance (τ_{00}), intercept/slope covariance (τ_{10}), slope variance (τ_{11}), TIC effect (γ_{02}), or time (γ_{10}).

I will describe the patterns of effects separately for each parameter. Average estimates, empirical standard deviations of the estimates, and relative bias for each parameter in the missing data conditions for Model 1 are given in Table 3 and Table 4, for 5 and 10 repeated measures respectively. The final numbers of retained replications (i.e., positive definite solutions and non-singular covariance matrices for imputation) for each cell in the design are also given in Tables 3 and 4. In Model 1, replications were only discarded in the ten repeated measures, 40% MCAR condition; 353 replications were discarded in this condition due to singular posterior covariance matrices for MCMC, which may cause imputed values of some variables to be fixed. RMSE for all Model 1 parameters are given in Table 5.

Intercept bias. For the intercept estimate (γ_{00}), there were main effects of the number of repeated measures and MI on bias. As in complete data, bias for the intercept decreased with more repeated measures; bias also decreased with the use of multiple imputation. For example, with 20% MCAR, relative bias drops from -14% (0.75 RMSE) with 5 repeated measures to -12% (0.62 RMSE) with 10 repeated measures. Similarly, for the 5 repeated measures, 20% MCAR condition, relative bias changes from -14% before to -13% after multiple imputation, from 0.75 to 0.70 RMSE. The main effect on relative bias for the number of repeated measures was more pronounced than the effect of MI. However, the main effect of MI does show that MI helps counteract intercept bias due to missing TVCs under these simulation conditions. Besides the main effects of the number of repeated measures and use of MI, no other factors or interactions among factors meaningfully predicted bias in the intercept parameter.

Between-person effect bias. For the estimate of the between-person effect (γ_{01}), the meta-model results showed main effects and an interaction between mechanism and repeated measures. Figure 1 displays the pattern among mechanism conditions. For a given number of repeated measures, relative bias is very similar in the MCAR and MAR conditions, about 76% for 5 repeated measures and about 45% for 10 repeated measures. However, with MNAR missingness, the average relative bias decreased to 49% and 8% for 5 and 10 repeated measures, respectively. This pattern is due to the exact MNAR process of the simulation design, such that MNAR missingness caused the between-person effect estimate to be underestimated, while at the same time bias in the person-mean caused the between-person effect to be overestimated. The main effect of repeated measures was consistent with predictions and the pattern in complete data, such that bias decreased with more repeated measures. For example, with 40% of TVCs MCAR and listwise deletion, relative bias decreased from 97.71 with 5 waves of data to 55.49 with 10. RMSE, reflecting variability and bias, decreased from 0.21 to 0.14. Interestingly, there was no main effect of MI, which suggests that MI does not meaningfully counteract bias in the between-person effect under these simulated conditions.

Within-person effect bias. There was an interaction between missingness mechanism and MI for the within-person TVC effect (γ_{20}) as well as main effects of mechanism and proportion missing. The nature of interaction is shown in Figure 2. The within-person TVC effect was generally estimated without bias before MI in all missingness conditions, however after MI this parameter exhibited some bias and the direction of the bias depended on the type of missingness. In the 10 repeated measures, 40% missing conditions, this parameter was estimated with relative bias less than 1 for MCAR, MAR, and MNAR missingness. However, after MI, this parameter was underestimated with relative bias of -4% for MCAR missingness and -6% for MAR missingness, but overestimated with 26% relative bias when the missingness was MNAR. This pattern corresponds to RMSE 0.03-0.04 before MI and 0.04-0.06 after MI. Figure 2 shows that the magnitude of this bias is fairly small relative to the size of the effect. The main effect of the proportion missing revealed that, in general, bias was greater with a higher proportion of cases missing. This is clear, for example, in the 10 repeated measures condition, after MI, where relative bias is 2% at 20% missing and 5.5% at 40% missing. This comparison is between modest RMSE values of 0.03 and 0.06.

Residual variance bias. The estimated residual variance (σ^2) was substantially affected by three factors: percent missing, mechanism, and MI. The three-way interaction between these factors was significant as well as all three two-way interactions and all three main effects. The interaction effects are illustrated in Figure 3. In general, this parameter did not show considerable bias, except after MI in the MNAR conditions. For instance, with 5 repeated measures and 20% MCAR or MAR missingness, relative bias did not exceed about 1%, before and after MI. However with 5 repeated measures and 20% MNAR missingness, relative bias jumped from 0% before to 14% after imputation. Not surprisingly, bias due to

MI of MNAR missingness was more pronounced at higher percent missing: relative bias jumped from less than 1% to 29% after MI in the 40% MNAR, 5 repeated measures condition. The RMSE for this parameter was 0.12 before MI for 20% MCAR, MAR, and MNAR missingness (5 repeated measures). After MI RMSE was essentially unchanged for MCAR and MAR missingness but rose to 0.32 for MNAR missingness.

Type I Errors. To detect potential spurious effects caused by TVC missingness, I examined the proportion of significant effects detected for an interaction between time and the between-person TVC effect (γ_{11}). Because this effect was zero in the population, the proportion of significant effects should be close to the chosen α level in a correctly specified model. Figure 4 shows the proportion of significant estimates, or Type I error rate, for this parameter across all conditions in Model 1, setting α = .05. In general, the Type I error rate stayed close to the nominal level, though in a few cases the rate is elevated after imputation.

Most noticeably, with 5 repeated measures and MNAR missingness, the Type I error rate is much higher after imputation. With 20% and 40% MNAR missing TVCs the Type 1 error rate reached 18% and 26% respectively after MI in the 5 repeated measures. Interestingly, the error rate was not inflated after MI in the 10 repeated measures conditions, staying at 7% and 8% for the 20 and 40% missing conditions, respectively.

Summary of Model 1 results. In sum, Model 1 results indicated that TVC missingness can substantially bias several model estimates, even if the missingness is completely at random. This bias is over and above bias found in complete data and generally due to increased sampling error in the estimate of the person mean of the TVC. MI restored bias to the complete data levels, unless the missingness was MNAR. Imputation of MNAR TVCs created new problems, including detection of a spurious interaction effect.

Model 2. Table 6 contains F-tests and partial η^2 values corresponding to each factor and interaction in the meta-models for Model 2. Mean estimates, standard deviations of the estimates, and relative bias for each parameter in the missing data conditions for Model 2 are given in Table 7 and Table 8, for five and 10 repeated measures respectively along with the number of retained replications within each cell. RMSE for all Model 2 parameters are summarized in Table 9.

The small random effect unique to Model 2 led to more non-positive definite solutions, especially with only five repeated measures. Before MI, the five repeated measures conditions lost an average of 43 replications, and after MI an average of 199 solutions were discarded. With 5 repeated measures, all discarded replications, before and after MI, were due to non-positive definite solutions. In the 10 repeated measures conditions, few replications were discarded due to non-positive definite solutions. The overall average number of replications retained was 964, but nearly all replications lost were in the 40% MCAR condition after MI. As in Model 1, this condition led to a high number of singular posterior covariance matrices which precluded MI, and 398 replications were discarded for this reason.

Meta-model results indicated that in Model 2, as in Model 1, there were nontrivial effects for the intercept (γ_{00}), between-person TVC effect (γ_{01}), and residual variance (σ^2). Effects were also found in Model 2 for the between (γ_{11}) and within (γ_{30}) person TVC effect interactions with time. Within the level-2 covariance matrix $\mathbf{T}$, there was only an effect on the variance of the random within-person effect (τ_{22}), and no effects on any of the other covariance parameters. There were no effects of any factor on the TIC (γ_{02}), time (γ_{10}), or within-person effect (γ_{20}).

Intercept bias. Just as in Model 1, there were main effects of the number of repeated measures and MI on bias in the intercept estimate such that bias for the intercept decreased with more repeated measures and with MI. Relative bias was -16% (0.82 RMSE) with 40% MCAR and 5 repeated measures before MI and -13% after (0.71 RMSE). Similarly with 10 repeated measures and 40% MCAR, relative bias was -13% before and -11% after MI, corresponding to 0.67 and 0.6 RMSE. Again, this suggests that MI decreases bias in the intercept estimate caused by missing TVCs.

Between-person effect bias. As in Model 1, main effects of mechanism and repeated measures predicted bias in the between-person effect (γ_{01}); additionally, in Model 2 there was an interaction between mechanism and MI. Again, bias decreased with more repeated measures. For example, in the MAR, 40% missing conditions (before MI) relative bias was 107% (0.23 RMSE) with 5 repeated measures, and with 10 repeated measures relative bias decreased to 73% (0.13 RMSE). As shown in Figure 5, MI decreased bias for MCAR and MAR but not MNAR missingness. One example of this can be seen in the 20% missing conditions with 10 repeated measures: relative bias decreases from 43.99 to 36.21 after MI with MCAR missingness and from 55.20 to 30.98 with MAR missingness, but in the MNAR condition the estimate is downwardly biased at -30.82 relative bias before MI and overestimed after MI with 39.77 relative bias.

Figure 5 also shows that the effect of listwise deletion of TVCs on the between-effect estimate differed considerably by mechanism. Relative bias was greatest with MAR missingness (91% with 20% missing, 5 repeated measures) and also quite high with MCAR missingness (79% with 20% missing, 5 repeated measures), however when the missingness was MNAR, bias was notably reduced (41% with 20% missing, 5 repeated measures). RMSE values for this parameter varied from 0.11-0.23 without MI to 0.10-0.15 with MI; these values are sizable considering the parameter value for the effect was 0.20.

Random within-person effect bias. Unlike in Model 1, no factors predicted bias in the within-person TVC effect (γ_{20}) in Model 2. However, the random effect variance for this parameter (τ_{22}), unique to Model 2, exhibited a meaningfully large main effect of MI such that the variance was underestimated after MI. Before MI, for example in Model 2 with 20% of TVCs MCAR and 5 repeated measures, relative bias for this estimate was trivial, at 1%, however after MI relative bias was -17%.

Bias in the time by between-person effect interaction. In Model 2 there was an interaction between mechanism and MI for raw bias in the estimate of the interaction of time with the between-person effect (γ_{11}), though this interaction also neared the medium effect-size threshold in the meta-model for Model 1. This parameter was zero in the population generating model and therefore relative bias is not defined for this estimate; however, in terms of raw bias, MI increased bias in the MNAR conditions. This can be seen in Figure 6 and in the mean estimates for the models with 5 repeated measures, where bias was minimal for all mechanisms, between -0.02 and 0.02 for all missingness conditions before MI but increased after MI, particularly for MNAR missingness. The raw bias increased to 0.06 and 0.10 for MNAR 20% and 40% missingness (5 repeated measures), which corresponded to RMSE values of 0.08 and 0.12.

Residual variance bias. The estimated residual variance was substantially affected by many factors and showed main effects of percent missing, mechanism, and MI as well as interactions between proportion missing and MI and between mechanism and MI. The interaction between proportion missing and MI, shown in Figure 7, was such that bias increased from the 20% to 40% missing conditions in the imputation conditions, but remained stable at essentially zero without MI. The second interaction, between MI and mechanism, is presented in Figure 8. This interaction shows that bias was negligible for all mechanisms before MI, but after MI bias increased slightly for MCAR and MAR conditions and substantially for MNAR conditions. Measuring variability and bias, RMSE values for this estimate varied considerably among the many conditions and was about 0.15 and 0.09 before MI for 5 and 10 repeated measures conditions, respectively. After MI, RMSE hovered around 0.16 for 5 repeated measures MCAR and MAR conditions, ranged between 0.11-0.22 for 10 repeated measures MCAR and MAR conditions, and jumped to 0.68 and 0.66 in the severe 40% missing 5 and 10 repeated measure MNAR conditions.

Bias in the time by within-person effect interaction. Bias in the estimated interaction between $\dot{z}_{ti}$ and time (γ_{30}) was explained by a main effect of mechanism and an interaction between mechanism and MI, plotted in Figure 9. Bias was negligible with and without MI for MCAR and MAR missingness conditions, but after MI the estimate was greatly overestimated in the MNAR conditions. Relative bias jumped from 12% to 177% with 40% of TVCs MNAR and 5 repeated measures; this was linked with an increase in RMSE from 0.04 to 0.10.

Type I Error Rates. In Model 2, the type I error rate for the interaction of time and the between-person effect showed some variation due to MI and mechanism. The rate was particularly high following MI with 5 repeated measures and MNAR missingness. The pattern of results for Model 2 was identical to the results described for Model 1. For simplicity I will not describe the pattern again here, but the rates can be seen in Figure 10.

For τ_{20} and τ_{21}, which were both zero in the population, the proportion of significant effects stayed close to or below the nominal desired alpha level in all conditions. This is evidence that Type I error rates for these estimates were not impacted by TVC missingness.

Summary of Model 2 results. As in Model 1, results in Model 2 indicated that TVC missingness can substantially bias model estimates, even if the missingness is completely at random. The amount and mechanism of missingness was also important, particularly because MI is not robust to non-ignorable missingness. However, bias due to missing TVCs must be considered relative to the bias known to occur in complete data due to sampling error of the person mean.

CHAPTER 4

DISCUSSION

I used a comprehensive simulation study to test my theoretically derived hypotheses concerning the effects of missing time-varying covariates in multilevel models. To my knowledge, this is the first study to specifically examine the effects of missing TVCs. The results of my project provided support for my research hypotheses and also revealed many complexities related to this type of missingness.

My results demonstrate that listwise deletion of missing TVCs can clearly bias model estimates; these effects should be understood and taken into consideration by researchers. Although TVC missingness is not ideal, the results of my project also highlight that the effects of TVC missingness may be overshadowed by bias already present even in complete data due to person-mean sampling error. My study illustrates that most bias due to TVC missingness occurs by increasing sampling error of the person mean. More repeated measures protected against this bias in complete and incomplete data. The mechanism of missingness was important because it directly determines if MI is appropriate and also because it systematically biased the person-mean estimate. In many cases MI counteracted bias due to missing TVCs, with some exceptions, particularly when TVCs were MNAR. I will briefly describe the effects of each factor in my simulation study.

Effects of Repeated Measures

As I predicted, more repeated measures generally protected against bias in model estimates. In both models, more repeated measures decreased the baseline level of bias in complete data. This result is consistent with what Lüdtke et al. (2008) predicted. When TVCs were missing, by any mechanism and in every condition in each model, bias was worse with five repeated measures than with 10. More repeated measures decreased bias in both the between-person effect estimate and the intercept estimate. For example, with 40% of TVCs MCAR and listwise deletion, relative bias in the between-person effect decreased from 97.71 with five waves of data to 55.49 with 10. It is apparent that bias in the between-person effect causes corresponding bias in the intercept. This is because the overall intercept estimate is forced downward to compensate for the overestimated between-person effect. With more repeated measures, a better estimate of the person mean is obtained, and the intercept estimate has less error for which to compensate.

These results can be generalized from the chosen set of population parameters to other population generating models. As shown in Equation 14, bias in the between-person effect will be in the direction of the true within-person effect. The intercept estimate will therefore be forced downward to accommodate upwardly biased between-person effects and vice versa, and the extent of the bias will of course depend on the magnitude of difference between the two effects, the ICC of the TVC, and the number of repeated measures.

Another important finding from both models in the simulation study was that the number of repeated measures influenced Type I error rates via the interaction between time and the between-person effect. A high Type I error rate was found in both models after imputation of MNAR missing TVCs in the five repeated measures conditions, but this bias was minimal with 10 repeated measures. It is not initially clear why this occurred. One distinctive feature of the high missingness MNAR conditions with five repeated measures is that a sizable proportion of replications end up with only one observation for the TVC (the first time point is always observed). With 10 repeated measures, cases with only one observation are much less likely, and it is possible the increased Type 1 error rate is related to imputation of this extreme, imbalanced missingness in the five repeated measures conditions.

Effects of Proportion Missing

I hypothesized that more missing TVCs would create more bias. There was some support for this hypothesis in the main effects of proportion missing found for the residual variance estimate and the within-person effect. Overall, however, the proportion missing was not a powerful explanatory factor in these simulations. The effect of proportion missing on the residual variance estimate was related to mechanism and MI in Model 1 and MI in Model 2. My results showed that bias in the residual variance estimate increased with the proportion missing after imputation in Model 2. This increase also occurred in Model 1, but only after imputation of MNAR missingness.

I did find that more missingness led to increased bias in the within-person effect, but only in Model 1. Because the within-person effect was generally estimated with high precision, bias due to increased missingness was noteworthy, but the magnitude of the bias was still small relative to the size of the effect, which had a population value of 1.00. For example in the 10 repeated measures, 40% missing conditions, the RMSE for this parameter was 0.03-0.04 before MI and 0.04-0.06 after MI. Surprisingly, increased missingness did not generally exaggerate bias in the between-person effect. This was likely due to the complicated patterns of more prominent effects due to other factors such as missingness mechanism and the use of multiple imputation.

Effects of Multiple Imputation and Missingness Mechanism

The mechanism of missingness was a key explanatory factor and was related to bias in several effects. As predicted, I found that the effect of mechanism frequently interacted with MI. Two major considerations are important in understanding these patterns of effects. First, MI is appropriate to correct for MCAR and MAR but not MNAR missingness. Secondly, the mechanism of TVC missingness differentially impacted the person mean estimate.

In some cases MI counteracted bias, but there were also some important exceptions. The use of MI offset bias due to missing TVCs in the intercept effect estimate, but after imputation the bias returned to the extremely high levels exhibited in complete data. Surprisingly, MI did not decrease bias in the between-person effect overall.

When missingness was MNAR, MI produced bias in the between- and within-person effects and interactions with these effects and time, as well as in the residual variance estimate. MI of MNAR missing TVCs also caused an increased Type I error rate of the interaction between the between-person effect and time. Statistical theory explains these effects, as it has been clearly established that MI is beneficial if missingness is at least MAR (Rubin, 1987). Unfortunately, if missingness is MNAR, we risk introducing bias by using MI. MI is not robust to the assumption of MAR missingness, though in applied research we must almost always depend on this assumption (see Enders, 2010; Schafer & Graham, 2002).

Notably, MI increased bias in the random within-person effect estimate, even when missingness was MCAR. For example in Model 2 with 20% of TVCs MCAR and five repeated measures, relative bias for this estimate was trivial when listwise deletion was used, however after MI relative bias was -17%. This important exception is a case where MI does not appropriately handle TVC missingness, and it is contrary to the general conclusion that

MI yields appropriate estimates for MAR and MCAR missingness (Rubin, 1987). Likely this finding is attributable to the considerable and widespread consequences of ignoring error in the person mean estimate. I believe the upwardly biased between-person effect may have obscured the more modest variance of the random within-person effect. This finding is especially interesting because it is not expected from missing data theory.

In sum, my simulation study clarified a number of key issues related to TVC missingness. The potential effects of TVC missingness should be considered accounting for the extent of bias that could be present in complete data due to sampling error in the person-mean, especially if few repeated measures are available. Multiple imputation of missing TVCs relies on the usual assumption of MAR missingness. Finally, due to the extensive effects of ignored bias in the between-person effect, multiple imputation may obscure some subtle but important estimates, such as the variance of the random within-person effect.

Implications and Recommendations for Applied Research

My findings have several implications for applied research. The results highlighted that TVC missingness can have unexpected effects and MI may not always protect against bias due to TVC missingness. My study is the first to demonstrate this and also to highlight the potentially extreme effects of sampling error in the person-mean. Given these results, I suspect that in many applications sampling error in the person-mean has likely affected inferences in published and unpublished research. The true between-person TVC effect has likely been obscured, particularly because growth models are often estimated with a small number of repeated measures. Currently, few researchers report details on missing TVCs, but this is an important potential problem for which researchers should have suitable solutions. My project also clearly demonstrated that the multilevel modeling framework has serious limitations for modeling between- and within-person TVC effects. A promising alternative is latent growth curve modeling (LCM) in the SEM framework which can allow for measurement error in the person mean to be accounted for (Curran et al., in press; Lüdtke et al., 2008). Based on the findings of this study, I propose three recommendations for applied researchers.

First, if a key research question concerns the between-person effect of a TVC, it should be a critical priority to decrease sampling error of the person-mean estimate. One way to do this is to collect more repeated assessments, but this option is rarely realistic. Instead, more reliable estimates of the TVC at each assessment may improve person-mean reliability; see Shrout and Lane (2012) for an excellent chapter on reliability. Equation 14 from Lüdtke et al. (2008) can be used to estimate bias in the between-person effect, which is also informative of bias in the intercept estimate.

Second, I recommend researchers use LCM instead of MLM to model bivariate relationships between a TVC and outcome. The key remaining advantage of MLM over LCM is flexibility to accommodate three-level nested data structures (Curran, 2003), and in the case of higher level nesting LCM may be intractable for this problem, despite potential bias in the between-person effect. Otherwise, I see no reason why researchers should not use the LCM framework in order to obtain an unbiased estimate of the between-person effect. See work by Curran et al. (in press) for details on disaggregating between- and within-person TVC effects in LCM.

Third, despite the most diligent efforts, some missing data is inevitable in longitudinal research. Therefore, in advance, researchers should collect data on potential correlates of missingness in an effort to satisfy the MAR assumption. As sampling error in the person mean decreases, holding other factors constant, the effects of missing TVCs will decrease. In the LCM framework, either multiple imputation or joint maximum likelihood estimation will prevent listwise deletion of TVCs. Both procedures assume an MAR missingness mechanism.

Limitations and Future Directions

The results from any simulation study can never be generalized to all cases. Parameter values were not varied in my simulation, but the majority of the findings would differ predictably given alternate sets of population parameters. Theoretically, TVCs could be missing an infinite number of ways, and there is no way to systematically study them all. Different forms of missing TVCs that do not systematically bias the mean would clearly not have the same effects on that estimate, however many of my results concerning missingness mechanism are generalizable. In this demonstration, the form of missingness was very strong and will almost certainly not be as strong in real data, though the extreme case presented here is still illustrative of the issue. The technique of MI I used may not be feasible for many researchers with less structured data. I do not believe this impairs the external validity of my results pertaining to MI, but a significant problem still remains to enable applied researchers to perform multilevel imputation using widely available software.

My study was limited to the standard multilevel modeling framework and did not consider strategies available in LCM. Another promising tool available in SEM is joint maximum likelihood estimation. With joint maximum likelihood, listwise deletion due to missing TVCs and the considerable complications of MI could be avoided. This has not been well-studied, but a future direction would be to compare joint maximum likelihood to multiple imputation of TVCs in LCM, accounting for measurement error in the person mean.

Conclusion

Missing TVCs are not ideal and listwise-deletion of missing TVCs may cause bias, but in many cases bias due to missing TVCs will be trivial compared to the bias that already exists in complete data due to error in the person mean. Potential bias due to missing TVCs will be less extreme with more repeated measures or if sampling error in the person-mean is otherwise reduced. MI can prevent loss of data but assumes MAR missingness at minimum. Missingness mechanism is impossible to determine in real data, and even if MNAR missingness is suspected, no general strategy exists to appropriately correct for MNAR missingness.

Currently, many researchers are estimating between- and within-person TVC effects in MLM, and my results suggest that there is a strong likelihood this often impedes our ability to draw accurate inferences. The more important issue highlighted in this project is that the multilevel modeling framework has serious limitations for modeling bivariate relationships over time. Alternative strategies for modeling these developmentally motivated relations are warranted. The SEM framework, because it is able to better address measurement error, is an alternative modeling framework that may be better suited for this type of analysis.

Table 1

Meta-model F-tests, Partial η^2 Values, and Average Parameter Estimates for Complete Data in Model 1 and Model 2

Model 1

Parameter (Pop Value)	Meta-model		5 Repeated Measures (R=1000)			10 Repeated Measures (R=1000)		
	F	(partial η^2)	Mean (SDEst)	Rel Bias	RMSE	Mean (SDEst)	Rel Bias	RMSE
Intercept, γ_{00}(5.0)	**145.57**	**(.07)**	4.34 (.22)	-13.28	.70	4.45 (.21)	-10.97	.59
TIC, γ_{02}(1.0)	0.13	(.00)	1.00 (.22)	.11	.22	1.00 (.21)	.44	.21
Time$_{ti}$, γ_{10}(.5)	0.19	(.00)	.50 (.12)	.31	.12	.50 (.11)	.75	.11
Time$_{ti}$ *$\bar{z}_i$, γ_{11}(0.0)	0.33	(.00)	.00 (.04)		.04	.00 (.05)		.05
$\bar{z}_i$, γ_{01}(.2)	**325.77**	**(.14)**	.33 (.07)	66.22	.15	.27 (.07)	36.01	.10
$\dot{z}_{ti}$, γ_{20}(1.0)	1.01	(.00)	1.00 (.04)	-.12	.04	1.00 (.02)	.02	.02
σ^2 (2.0)	0.28	(.00)	2.00 (.10)	.08	.10	2.00 (.10)	.18	.06
τ_{00} (2.0)	51.62	(.03)	2.18 (.29)	8.91	.34	2.18 (.29)	4.54	.26
τ_{10} (-0.2)	0.27	(.00)	-.20 (.14)	-2.46	.14	-.20 (.14)	-1.02	.11
τ_{11} (1.0)	5.13	(.00)	1.00 (.10)	.47	.10	1.00 (.10)	1.46	.09

Model 2

	Meta-Model		5 Repeated Measures (R=993)			10 Repeated Measures (R=1000)		
	F	(partial η^2)	Mean (SDEst)	Rel Bias	RMSE	Mean (SDEst)	Rel Bias	RMSE
Intercept, γ_{00} (5.0)	**131.65**	**(.06)**	4.34 (.22)	-13.24	.70	4.45 (.21)	-11.02	.59
TIC, γ_{02} (1.0)	0.18	(.00)	1.00 (.22)	.08	.22	1.00 (.21)	.49	.21
Time$_{ti}$, γ_{10} (0.5)	4.90	(.00)	.48 (.12)	-3.21	.12	.50 (.11)	-.97	.11
Time$_{ti}$ *$\bar{z}_i$, γ_{11} (0.0)	2.67	(.00)	.01 (.05)		.05	.00 (.05)		.05
Time$_{ti}$ *$\dot{z}_{ti}$, γ_{30} (.05)	0.05	(.00)	.05 (.03)	.15	.03	.05 (.01)	.61	.01
$\bar{z}_i$, γ_{01} (0.2)	**294.40**	**(.13)**	.33 (.08)	65.62	.15	.27 (.08)	36.43	.11
$\dot{z}_{ti}$, γ_{20} (1.0)	0.24	(.00)	1.00 (.07)	-.22	.07	1.00 (.05)	-.08	.05
σ^2 (2.0)	0.33	(.00)	2.00 (.12)	.05	.12	2.00 (.07)	.17	.07
τ_{00} (2.0)	70.21	(.03)	2.21 (.31)	10.74	.38	2.11 (.25)	5.46	.27
τ_{10} (-0.2)	2.00	(.00)	-.18 (.14)	-9.00	.14	-.19 (.11)	-5.02	.11
τ_{11} (1.0)	4.27	(.00)	1.01 (.11)	.57	.11	1.01 (.09)	1.48	.09
τ_{20} (0.0)	0.49	(.00)	.00 (.07)		.07	.00 (.05)		.05
τ_{21} (0.0)	0.25	(.00)	.00 (.05)		.05	.00 (.03)		.03
τ_{22} (0.1)	0.18	(.00)	.10 (.04)	-1.01	.04	.10 (.02)	-1.57	.02

Note. R = replications per condition. The F-test degrees of freedom are (1,1998) in Model 1 and (1,1991) in Model 2.

Table 2

Model 1 Meta-model F-tests and Partial η^2 Values for Each Simulation Factor in Missing Data Conditions

Factor (*df*)	Intercept (γ_{00}) F (partial η^2)	TIC (γ_{02}) F (partial η^2)	Time$_{ti}$ (γ_{10}) F (partial η^2)	Time$_{ti}$ $*\bar{z}_i$ (γ_{11}) F (partial η^2)	$\bar{z}_i$ (γ_{01}) F (partial η^2)	$\dot{z}_{ti}$ (γ_{20}) F (partial η^2)
prctmiss (1)	631.30 (.03)	0.08 (.00)	40.33 (.00)	33.22 (.00)	275.89 (.01)	0.00 (.03)
mech (2)	354.71 (.03)	0.15 (.00)	36.69 (.00)	231.93 (.02)	**1413.13 (.11)**	**0.00 (.08)**
rms (1)	**2063.71 (.08)**	1.72 (.00)	10.77 (.00)	154.54 (.01)	**3779.81 (.14)**	4.01 (.00)
MI (1)	**1689.15 (.07)**	0.06 (.00)	1.81 (.00)	1138.50 (.05)	1004.53 (.04)	528.06 (.02)
*prctmiss*mech* (2)	250.86 (.02)	0.04 (.00)	27.40 (.00)	8.99 (.00)	50.12 (.00)	14.40 (.00)
*prctmiss*rms* (1)	0.21 (.00)	0.01 (.00)	16.52 (.00)	13.65 (.00)	1.80 (.00)	1.16 (.00)
*prctmiss*MI* (1)	425.08 (.02)	0.00 (.00)	14.21 (.00)	134.85 (.01)	739.81 (.03)	721.37 (.03)
*mech*rms* (2)	4.82 (.00)	0.19 (.00)	15.78 (.00)	87.63 (.01)	37.03 (.00)	115.11 (.01)
*mech*MI* (2)	225.38 (.02)	0.02 (.00)	3.51 (.00)	646.60 (.05)	574.87 (.05)	**1171.72 (.09)**
*rms*MI* (1)	3.15 (.00)	0.01 (.00)	3.87 (.00)	261.20 (.01)	582.37 (.02)	0.00 (.00)
*prctmiss*mech*rms* (2)	3.01 (.00)	0.01 (.00)	6.69 (.00)	5.77 (.00)	4.52 (.00)	23.67 (.00)
*prctmiss*mech*MI* (2)	147.10 (.01)	0.03 (.00)	8.72 (.00)	70.86 (.01)	93.84 (.01)	8.24 (.00)
*prctmiss*rms*MI* (1)	11.63 (.00)	0.01 (.00)	6.78 (.00)	20.53 (.00)	7.78 (.00)	3.15 (.00)
*mech*rms*MI* (2)	4.23 (.00)	0.29 (.00)	6.52 (.00)	124.05 (.01)	219.92 (.02)	109.14 (.01)
*prctmiss*mech*rms*MI* (2)	13.88 (.00)	0.12 (.00)	6.19 (.00)	6.46 (.00)	0.50 (.00)	18.92 (.00)
	σ^2	τ_{00}	τ_{10}	τ_{11}		
prctmiss (1)	**1772.05 (.07)**	10.43 (.00)	0.03 (.00)	0.33 (.00)		
mech (2)	**7063.25 (.37)**	13.23 (.00)	7.33 (.00)	1.22 (.00)		
rms (1)	11.31 (.00)	443.24 (.02)	8.70 (.00)	62.04 (.00)		
MI (1)	**13465.90 (.36)**	888.00 (.04)	12.41 (.00)	0.00 (.00)		
*prctmiss*mech* (2)	**746.95 (.06)**	3.58 (.00)	1.83 (.00)	0.41 (.00)		
*prctmiss*rms* (1)	4.89 (.00)	1.51 (.00)	0.43 (.00)	0.07 (.00)		
*prctmiss*MI* (1)	**1816.82 (.07)**	137.26 (.01)	5.64 (.00)	0.18 (.00)		
*mech*rms* (2)	62.07 (.01)	21.77 (.00)	5.13 (.00)	1.06 (.00)		
*mech*MI* (2)	**6973.36 (.37)**	180.13 (.02)	10.26 (.00)	0.75 (.00)		
*rms*MI* (1)	3.11 (.00)	48.07 (.00)	8.75 (.00)	0.11 (.00)		
*prctmiss*mech*rms* (2)	13.59 (.00)	4.77 (.00)	0.70 (.00)	0.11 (.00)		
*prctmiss*mech*MI* (2)	**706.84 (.06)**	20.91 (.00)	3.36 (.00)	0.37 (.00)		
*prctmiss*rms*MI* (1)	2.32 (.00)	4.40 (.00)	2.74 (.00)	0.05 (.00)		
*mech*rms*MI* (2)	57.47 (.00)	12.18 (.00)	4.50 (.00)	0.27 (.00)		
*prctmiss*mech*rms*MI* (2)	7.87 (.00)	0.57 (.00)	0.91 (.00)	0.39 (.00)		

Note. Medium-size partial η^2 value of .06 or larger are shown in bold.

Table 3

Average Parameter Estimates for Model 1 Missing Data Conditions, 5 Repeated Measures

	Listwise											
	MCAR				MAR				MNAR			
	20%		40%		20%		40%		20%		40%	
Parameter (Pop Value)	Mean (SD)	Rel Bias	Mean (SD)	Rel Bias	Mean (SD)	Rel Bias	Mean (SD)	Rel Bias	Mean (SD)	Rel Bias	Mean (SD)	Rel Bias
Intercept, $\gamma_{00}(5.0)$	4.28 (.22)	-14.33	4.21 (.22)	-15.76	4.28 (.22)	-14.41	4.24 (.22)	-15.18	4.24 (.22)	-15.29	3.99 (.21)	-20.25
TIC, γ_{02} (1.0)	1.00 (.22)	-.19	1.00 (.23)	-.50	1.00 (.22)	.06	1.00 (.23)	.11	1.00 (.22)	.02	1.00 (.23)	.05
Time_{ti} , $\gamma_{10}(.5)$	.50 (.12)	.31	.50 (.12)	.82	.50 (.12)	.26	.50 (.12)	.74	.51 (.12)	1.75	.52 (.11)	4.38
$\text{Time}_{ti} * \bar{z}_{\text{i}}$, γ_{11} (0.0)	.00 (.05)		.00 (.05)		.00 (.05)		.00 (.05)		-.01 (.06)		-.02 (.07)	
$\bar{z}_{\text{i}}$, $\gamma_{01}(.2)$	.36 (.08)	79.48	.40 (.07)	97.71	.38 (.07)	91.69	.41 (.07)	107.04	.28 (.09)	40.01	.37 (.09)	83.52
$\dot{z}_{ti}$, $\gamma_{20}(1.0)$	1.00 (.04)	-.04	1.00 (.05)	-.05	1.00 (.04)	.00	1.00 (.05)	-.03	1.00 (.05)	-.07	1.00 (.06)	-.48
σ^2 (2.0)	2.00 (.12)	.07	2.00 (.15)	-.08	2.00 (.12)	.22	2.00 (.14)	-.04	2.00 (.12)	-.06	2.00 (.14)	.21
τ_{00} (2.0)	2.22 (.31)	11.03	2.27 (.34)	13.72	2.22 (.32)	10.89	2.28 (.35)	13.80	2.26 (.32)	13.07	2.33 (.35)	16.71
τ_{10} (-0.2)	-.20 (.14)	2.28	-.20 (.16)	-2.14	-.20 (.15)	-1.82	-.20 (.16)	1.98	-.20 (.15)	-2.35	-.20 (.17)	2.11
τ_{11} (1.0)	1.00 (.11)	.45	1.00 (.12)	.35	1.00 (.11)	.45	1.00 (.12)	.41	1.00 (.11)	.23	1.00 (.12)	.29
	Multiple Imputation											
	MCAR				MAR				MNAR			
	20%		40%		20%		40%		20%		40%	
Intercept, $\gamma_{00}(5.0)$	4.33 (.22)	-13.30	4.33 (.23)	-13.32	4.35 (.22)	-13.07	4.34 (.23)	-13.23	4.34 (.22)	-13.28	4.28 (.22)	-14.37
TIC, γ_{02} (1.0)	1.00 (.22)	-.01	1.00 (.22)	-.03	1.00 (.22)	.07	1.00 (.22)	.05	1.00 (.22)	-.12	1.00 (.22)	-.14
Time_{ti} , $\gamma_{10}(.5)$	.50 (.12)	.52	.50 (.12)	.94	.49 (.12)	-1.11	.50 (.12)	.66	.50 (.12)	.16	.57 (.12)	13.04
$\text{Time}_{ti} * \bar{z}_{\text{i}}$, γ_{11} (0.0)	.00 (.05)		.00 (.05)		.01 (.05)		.03 (.05)		.05 (.06)		.09 (.07)	
$\bar{z}_{\text{i}}$, $\gamma_{01}(.2)$	.33 (.08)	66.47	.33 (.08)	66.37	.32 (.08)	57.51	.29 (.09)	45.70	.29 (.10)	43.20	.26 (.12)	30.03
$\dot{z}_{ti}$, $\gamma_{20}(1.0)$	.99 (.04)	-.86	.98 (.04)	-2.22	.98 (.04)	-1.78	.95 (.04)	-4.59	1.03 (.05)	2.85	.99 (.06)	-1.08
σ^2 (2.0)	2.02 (.12)	.79	2.04 (.14)	2.15	2.02 (.12)	1.23	2.06 (.13)	2.92	2.29 (.13)	14.33	2.58 (.18)	28.97
τ_{00} (2.0)	2.18 (.30)	8.76	2.16 (.32)	8.13	2.17 (.31)	8.72	2.17 (.32)	8.32	2.08 (.31)	4.12	1.97 (.34)	-1.74
τ_{10} (-0.2)	-.19 (.14)	-2.69	-.19 (.15)	-3.97	-.20 (.14)	-1.28	-.20 (.15)	-.48	-.18 (.14)	-8.91	-.16 (.15)	-21.13
τ_{11} (1.0)	1.00 (.11)	.40	1.00 (.11)	.33	1.01 (.11)	.72	1.01 (.11)	.94	1.00 (.11)	.41	1.00 (.12)	-.34

Note. There were 1000 replications per condition.

Table 4

Average Parameter Estimates for Model 1 Missing Data Conditions, 10 Repeated Measures

Parameter (Pop Value)	Listwise											
	MCAR				MAR				MNAR			
	20% (*R*=1000)		40% (*R*=1000)		20% (*R*=1000)		40% (*R*=1000)		20% (*R*=1000)		40% (*R*=1000)	
	Mean	Rel Bias	Mean	Rel Bias	Mean	Rel Bias	Mean	Rel Bias	Mean	Rel Bias	Mean	Rel Bias
Intercept, γ_{00}(5.0)	4.42 (.22)	-11.59	4.37 (.23)	-12.53	4.41 (.21)	-11.88	4.36 (.21)	-12.86	4.41 (.22)	-11.83	4.08 (.20)	-18.40
TIC, γ_{02} (1.0)	1.00 (.21)	.45	1.01 (.22)	.57	1.00 (.21)	.38	1.00 (.22)	.39	1.00 (.22)	.12	1.00 (.22)	-.26
Time$_{ti}$, γ_{10}(.5)	.50 (.11)	.78	.50 (.11)	.39	.50 (.11)	.65	.50 (.11)	.42	.51 (.11)	1.01	.51 (.10)	2.64
Time$_{ti}$ *$\bar{z}_i$, γ_{11} (0.0)	.00 (.05)		.00 (.05)		.00 (.04)		.00 (.05)		.00 (.06)		-.02 (.07)	
$\bar{z}_i$, γ_{01}(.2)	.29 (.08)	43.72	.31 (.08)	55.49	.31 (.08)	55.09	.35 (.08)	72.65	.14 (.10)	-30.98	.22 (.11)	11.84
$\dot{z}_{ti}$, γ_{20}(1.0)	1.00 (.03)	.05	1.00 (.03)	.02	1.00 (.03)	.06	1.00 (.03)	.10	1.00 (.03)	-.02	1.00 (.04)	-.27
σ^2 (2.0)	2.00 (.12)	.14	2.00 (.15)	.14	2.00 (.12)	.20	2.00 (.14)	.25	2.00 (.12)	.21	2.00 (.14)	.20
τ_{00} (2.0)	2.22 (.31)	5.69	2.27 (.34)	7.19	2.22 (.32)	5.77	2.28 (.35)	8.10	2.26 (.32)	7.69	2.33 (.35)	12.08
τ_{10} (-0.2)	-.20 (.14)	-1.31	-.20 (.16)	-2.41	-.20 (.15)	-1.47	-.20 (.16)	.48	-.20 (.15)	-1.07	-.20 (.17)	1.25
τ_{11} (1.0)	1.00 (.11)	1.44	1.00 (.12)	1.44	1.00 (.11)	1.45	1.00 (.12)	1.49	1.00 (.11)	1.57	1.00 (.12)	1.49
	Multiple Imputation											
	MCAR				MAR				MNAR			
	20% (*R*=1000)		40% (*R*=647)		20% (*R*=1000)		40% (*R*=1000)		20% (*R*=1000)		40% (*R*=1000)	
Intercept, γ_{00}(5.0)	4.45 (.22)	-11.00	4.45 (.22)	-10.92	4.46 (.22)	-10.85	4.45 (.22)	-10.92	4.46 (.22)	-10.84	4.44 (.21)	-11.12
TIC, γ_{02} (1.0)	1.01 (.21)	.58	1.00 (.21)	.11	1.00 (.21)	.38	1.00 (.21)	.43	1.00 (.21)	.35	1.00 (.21)	.36
Time$_{ti}$, γ_{10}(.5)	.50 (.11)	.88	.50 (.11)	.93	.50 (.11)	.25	.50 (.11)	.41	.50 (.11)	-.29	.51 (.11)	2.60
Time$_{ti}$*$\bar{z}_i$, γ_{11} (0.0)	.00 (.05)	.	.00 (.05)	.	.00 (.05)	.	.01 (.05)	.	.02 (.06)	.	.03 (.07)	.
$\bar{z}_i$, γ_{01}(.2)	.27 (.08)	36.15	.27 (.08)	35.78	.26 (.08)	31.91	.25 (.09)	25.96	.26 (.09)	28.48	.25 (.11)	23.78
$\dot{z}_{ti}$, γ_{20}(1.0)	.99 (.02)	-1.35	.96 (.03)	-4.37	.98 (.02)	-2.00	.94 (.03)	-5.54	1.04 (.03)	4.31	1.02 (.04)	1.88
σ^2 (2.0)	2.03 (.07)	1.58	2.09 (.09)	4.52	2.04 (.07)	1.95	2.10 (.08)	5.02	2.26 (.08)	13.13	2.53 (.11)	26.44
τ_{00} (2.0)	2.09 (.25)	4.34	2.07 (.27)	3.61	2.09 (.26)	4.32	2.09 (.27)	4.36	2.05 (.26)	2.74	2.01 (.27)	.43
τ_{10} (-0.2)	-.20 (.11)	-1.41	-.20 (.11)	-1.87	-.20 (.11)	-.75	-.20 (.11)	1.09	-.20 (.11)	-1.29	-.19 (.12)	-3.27
τ_{11} (1.0)	1.01 (.09)	1.43	1.01 (.09)	1.23	1.01 (.09)	1.49	1.02 (.09)	1.58	1.01 (.09)	1.50	1.01 (.09)	1.40

Note. R=Replications per condition.

Table 5

Model 1 RMSE

	Listwise													
	5 Repeated Measures							10 Repeated Measures						
		MCAR		MAR		MNAR			MCAR		MAR		MNAR	
Parameter (Pop Value)	Complete	20%	40%	20%	40%	20%	40%	Complete	20%	40%	20%	40%	20%	40%
Intercept, γ_{00}(5.0)	.70	.75	.82	.75	.79	.80	1.03	.59	.62	.67	.63	.68	.63	.94
TIC, γ_{02} (1.0)	.22	.22	.23	.22	.22	.22	.23	.21	.21	.22	.21	.22	.22	.22
Time$_{ti}$, γ_{10}(.5)	.12	.12	.12	.12	.12	.12	.12	.11	.11	.11	.11	.11	.11	.10
Time$_{ti}$ *$\bar{z}_i$, γ_{11} (0.0)	.04	.05	.05	.05	.05	.06	.07	.05	.05	.05	.04	.05	.06	.07
$\bar{z}_i$, γ_{01}(.2)	.15	.18	.21	.20	.23	.12	.19	.10	.12	.14	.13	.16	.12	.11
$\dot{z}_{ti}$, γ_{20}(1.0)	.04	.04	.05	.04	.05	.05	.06	.02	.03	.03	.03	.03	.03	.04
σ^2 (2.0)	.10	.12	.15	.12	.14	.12	.14	.06	.07	.09	.07	.09	.07	.09
τ_{00} (2.0)	.34	.38	.43	.38	.44	.41	.48	.26	.28	.32	.29	.33	.31	.38
τ_{10} (-0.2)	.14	.14	.16	.15	.16	.15	.17	.11	.11	.12	.11	.12	.12	.13
τ_{11} (1.0)	.10	.11	.12	.11	.12	.11	.12	.09	.09	.09	.09	.10	.09	.10
	Multiple Imputation													
	5 Repeated Measures							10 Repeated Measures						
		MCAR		MAR		MNAR			MCAR		MAR		MNAR	
	Complete	20%	40%	20%	40%	20%	40%	Complete	20%	40%	20%	40%	20%	40%
Intercept, γ_{00}(5.0)	.70	.70	.70	.69	.70	.70	.75	.59	.59	.59	.58	.59	.58	.59
TIC, γ_{02} (1.0)	.22	.22	.22	.22	.22	.22	.22	.21	.21	.21	.21	.21	.21	.21
Time$_{ti}$, γ_{10}(.5)	.12	.12	.12	.12	.12	.12	.13	.11	.11	.11	.11	.11	.11	.11
Time$_{ti}$ *$\bar{z}_i$, γ_{11} (0.0)	.04	.05	.05	.05	.06	.08	.11	.05	.05	.05	.05	.05	.06	.07
$\bar{z}_i$, γ_{01}(.2)	.15	.15	.15	.14	.13	.13	.13	.10	.11	.11	.10	.10	.11	.12
$\dot{z}_{ti}$, γ_{20}(1.0)	.04	.04	.05	.04	.06	.06	.06	.02	.03	.05	.03	.06	.05	.04
σ^2 (2.0)	.10	.12	.14	.12	.14	.32	.61	.06	.08	.12	.08	.13	.27	.54
τ_{00} (2.0)	.34	.35	.36	.35	.36	.32	.34	.26	.27	.28	.27	.28	.26	.27
τ_{10} (-0.2)	.14	.14	.15	.14	.15	.14	.16	.11	.11	.11	.11	.11	.11	.12
τ_{11} (1.0)	.10	.11	.11	.11	.11	.11	.12	.09	.09	.09	.09	.09	.09	.09

Table 6

Model 2 Meta-model F-tests and Partial η^2 Values for Each Simulation Factor in Missing Data Conditions

Factor (*df*)	Intercept (γ_{00}) F (partial η^2)	TIC (γ_{02}) F (partial η^2)	Time$_{ti}$ (γ_{10}) F (partial η^2)	Time$_{ti}$ *$\bar{z}_i$ (γ_{11}) F (partial η^2)	Time$_{ti}$ *$\dot{z}_{ti}$ (γ_{30}) F (partial η^2)	$\bar{z}_i$ (γ_{01}) F (partial η^2)	$\dot{z}_{ti}$ (γ_{20}) F (partial η^2)
prctmiss (1)	520.16 (.02)	0.40 (.00)	14.18 (.00)	51.32 (.00)	273.65 (.01)	310.33 (.01)	653.21 (.03)
mech (2)	282.63 (.02)	0.45 (.00)	21.73 (.00)	107.38 (.01)	**1862.88 (.14)**	**760.28 (.06)**	191.97 (.02)
rms (1)	**1837.24 (.08)**	3.79 (.00)	4.13 (.00)	437.01 (.02)	350.89 (.02)	**3256.89 (.13)**	47.98 (.00)
_MI_ (1)	**1610.88 (.07)**	0.03 (.00)	18.22 (.00)	1071.91 (.05)	1206.43 (.05)	609.88 (.03)	1200.30 (.05)
prctmiss*mech (2)	183.36 (.02)	0.12 (.00)	12.75 (.00)	10.08 (.00)	339.34 (.03)	82.54 (.01)	188.94 (.02)
prctmiss*rms (1)	0.13 (.00)	0.32 (.00)	19.23 (.00)	20.69 (.00)	63.46 (.00)	1.25 (.00)	1.06 (.00)
prctmiss*MI (1)	423.26 (.02)	0.01 (.00)	27.50 (.00)	114.10 (.01)	213.29 (.01)	606.37 (.03)	540.52 (.02)
mech*rms (2)	4.91 (.00)	0.05 (.00)	18.40 (.00)	120.91 (.01)	360.83 (.03)	23.32 (.00)	90.89 (.01)
mech*MI (2)	225.83 (.02)	0.04 (.00)	7.02 (.00)	**728.60 (.06)**	**1474.68 (.12)**	**871.98 (.07)**	105.96 (.01)
rms*MI (1)	2.74 (.00)	0.34 (.00)	4.35 (.00)	231.38 (.01)	294.50 (.01)	579.35 (.03)	29.38 (.00)
prctmiss*mech*rms (2)	1.92 (.00)	0.01 (.00)	8.31 (.00)	8.11 (.00)	55.98 (.01)	2.55 (.00)	6.85 (.00)
prctmiss*mech*MI (2)	157.02 (.01)	0.08 (.00)	13.62 (.00)	68.34 (.01)	238.31 (.02)	54.34 (.00)	117.86 (.01)
prctmiss*rms*MI (1)	10.67 (.00)	0.16 (.00)	5.65 (.00)	21.82 (.00)	53.69 (.00)	6.00 (.00)	1.21 (.00)
mech*rms*MI (2)	3.71 (.00)	0.18 (.00)	7.25 (.00)	105.39 (.01)	260.01 (.02)	233.04 (.02)	73.12 (.01)
prctmiss*mech*rms*MI (2)	13.16 (.00)	0.12 (.00)	5.25 (.00)	6.73 (.00)	36.94 (.00)	0.47 (.00)	5.73 (.00)
	σ^2	τ_{00}	τ_{10}	τ_{11}	τ_{20}	τ_{21}	τ_{22}
prctmiss (1)	2533.95 (.10)	30.14 (.00)	0.00 (.00)	0.01 (.00)	67.37 (.00)	2.58 (.00)	158.48 (.00)
mech (2)	**5776.68 (.33)**	4.16 (.00)	9.37 (.00)	1.99 (.00)	352.83 (.03)	2.28 (.00)	229.57 (.02)
rms (1)	120.45 (.01)	686.03 (.03)	23.41 (.00)	45.37 (.00)	5.78 (.00)	0.26 (.00)	152.09 (.02)
_MI_ (1)	**21188.40 (.48)**	1301.15 (.05)	1.86 (.00)	0.45 (.00)	19.76 (.00)	0.60 (.00)	**2787.06 (.08)**
prctmiss*mech (2)	634.29 (.05)	1.24 (.00)	1.67 (.00)	0.41 (.00)	52.06 (.00)	3.73 (.00)	13.80 (.00)
prctmiss*rms (1)	35.69 (.00)	0.00 (.00)	0.03 (.00)	0.00 (.00)	1.54 (.00)	3.27 (.00)	19.16 (.00)
prctmiss*MI (1)	**2760.06 (.11)**	202.85 (.01)	2.51 (.00)	0.01 (.00)	0.33 (.00)	0.17 (.00)	283.91 (.01)
mech*rms (2)	51.16 (.00)	16.15 (.00)	3.59 (.00)	1.22 (.00)	7.12 (.00)	5.85 (.00)	9.22 (.00)
mech*MI (2)	**5816.18 (.33)**	202.13 (.02)	7.78 (.00)	1.38 (.00)	52.85 (.00)	0.79 (.00)	148.65 (.02)
rms*MI (1)	26.24 (.00)	60.71 (.00)	8.75 (.00)	1.15 (.00)	23.75 (.00)	1.16 (.00)	1.15 (.00)
prctmiss*mech*rms (2)	5.87 (.00)	2.14 (.00)	0.09 (.00)	0.01 (.00)	3.29 (.00)	2.88 (.00)	1.81 (.00)
prctmiss*mech*MI (2)	625.46 (.05)	21.19 (.00)	2.08 (.00)	0.35 (.00)	1.93 (.00)	0.57 (.00)	4.84 (.00)
prctmiss*rms*MI (1)	11.34 (.00)	5.30 (.00)	2.87 (.00)	0.02 (.00)	2.46 (.00)	0.30 (.00)	3.05 (.00)
mech*rms*MI (2)	59.64 (.00)	10.32 (.00)	5.57 (.00)	0.47 (.00)	8.85 (.00)	0.15 (.00)	0.10 (.00)
prctmiss*mech*rms*MI (2)	4.94 (.00)	0.13 (.00)	0.59 (.00)	0.35 (.00)	1.26 (.00)	0.08 (.00)	0.60 (.00)

Note. Medium-size partial η^2 value of .06 or larger are shown in bold.

Table 7

Average Parameter Estimates for Model 2 Missing Data Conditions, 5 Repeated Measures

	Listwise											
	MCAR				MAR				MNAR			
	20 (*R*=979)		40 (*R*=955)		20 (*R*=988)		40 (*R*=952)		20 (*R*=964)		40 (*R*=902)	
Parameter (Pop Value)	Mean	Rel Bias	Mean	Rel Bias	Mean	Rel Bias	Mean	Rel Bias	Mean	Rel Bias	Mean	Rel Bias
Intercept, γ_{00}(5.0)	4.28 (.23)	-14.36	4.21 (.23)	-15.77	4.28 (.22)	-14.39	4.24 (.22)	-15.20	4.23 (.22)	-15.34	3.99 (.21)	-20.29
TIC, γ_{02} (1.0)	1.00 (.23)	.02	1.00 (.23)	-.43	1.00 (.23)	.02	1.00 (.23)	.08	1.00 (.22)	.03	1.00 (.23)	-.40
Time_{ti} , γ_{10}(.5)	.48 (.12)	-3.72	.48 (.12)	-3.76	.48 (.12)	-3.87	.48 (.12)	-3.79	.49 (.12)	-2.32	0.49 (.12)	-1.86
$\text{Time}_{ti}*\bar{z}_{\text{i}}$, γ_{11} (0.0)	.01 (.05)		.01 (.05)		.01 (.05)		.02 (.05)		.00 (.06)		-0.02 (.07)	
$\text{Time}_{ti}*\dot{z}_{ti}$, γ_{30}(.05)	.05 (.03)	.86	.05 (.04)	.50	.05 (.03)	-.45	.05 (.04)	-2.85	.05 (.04)	3.90	0.06 (.04)	12.30
$\bar{z}_{\text{i}}$, γ_{01}(.2)	.36 (.08)	79.27	.39 (.08)	97.42	.38 (.08)	91.42	.41 (.08)	107.21	.28 (.09)	40.59	0.37 (.09)	84.66
$\dot{z}_{ti}$, γ_{20}(1.0)	1.00 (.08)	-.18	1.00 (.09)	-.16	1.00 (.08)	-.13	1.00 (.09)	.07	.99 (.08)	-.51	0.99 (.09)	-1.39
σ^2 (2.0)	2.00 (.14)	-.18	1.99 (.17)	-.47	2.00 (.14)	.15	1.99 (.16)	-.39	2.00 (.14)	-.15	1.99 (.16)	-.50
τ_{00} (2.0)	2.27 (.33)	13.44	2.34 (.36)	17.02	2.27 (.33)	13.43	2.35 (.37)	17.50	2.32 (.33)	16.07	2.44 (.38)	22.12
τ_{10} (-0.2)	-.18 (.15)	-9.89	-.18 (.16)	-11.28	-.18 (.15)	-9.05	-.19 (.17)	-4.64	-.18 (.15)	-10.46	-.19 (.18)	-4.08
τ_{11} (1.0)	1.01 (.11)	.57	1.00 (.12)	.35	1.01 (.11)	.58	1.00 (.13)	.46	1.00 (.11)	.26	1.01 (.13)	.54
τ_{20} (0.0)	.00 (.09)		.00 (.10)		.00 (.09)		.00 (.10)		-.02 (.09)		-.05 (.12)	
τ_{21} (0.0)	.00 (.05)		.00 (.06)		.00 (.05)		.00 (.06)		.00 (.06)		.00 (.07)	
τ_{22} (0.10)	.10 (.05)	1.18	.10 (.05)	3.48	.10 (.04)	-.58	.10 (.05)	3.29	.10 (.05)	1.85	.11 (.06)	9.89
	Multiple Imputation											
	20 (*R*=937)		40 (*R*=765)		20 (*R*=910)		40 (*R*=748)		20 (*R*=877)		40 (*R*=570)	
Intercept, γ_{00}(5.0)	4.33 (.23)	-13.34	4.33 (.23)	-13.39	4.35 (.23)	-13.02	4.34 (.23)	-13.18	4.33 (.23)	-13.33	4.30 (.22)	-14.01
TIC, γ_{02} (1.0)	1.00 (.22)	-.03	1.00 (.23)	-.11	1.00 (.22)	.25	1.00 (.23)	-.48	1.00 (.22)	-.21	.99 (.22)	-.86
Time_{ti} , γ_{10}(.5)	.49 (.12)	-2.79	.49 (.12)	-2.99	.47 (.12)	-5.22	.49 (.12)	-2.90	.48 (.12)	-3.14	.55 (.12)	9.60
$\text{Time}_{ti}*\bar{z}_{\text{i}}$, γ_{11} (0.0)	.01 (.05)		.01 (.05)		.02 (.05)		.04 (.05)		.06 (.06)		.10 (.07)	
$\text{Time}_{ti}*\dot{z}_{ti}$, γ_{30}(.05)	.05 (.03)	-6.07	.04 (.04)	-11.63	.05 (.03)	2.31	.05 (.04)	6.00	.09 (.04)	73.89	.14 (.04)	177.00
$\bar{z}_{\text{i}}$, γ_{01}(.2)	.33 (.08)	66.70	.33 (.08)	66.18	.31 (.08)	56.02	.29 (.09)	43.11	.30 (.10)	49.66	.29 (.12)	44.35
$\dot{z}_{ti}$, γ_{20}(1.0)	.99 (.08)	-.52	.99 (.08)	-1.44	.98 (.08)	-2.28	.94 (.09)	-5.66	.97 (.08)	-2.71	.87 (.09)	-13.43
σ^2 (2.0)	2.05 (.13)	2.48	2.11 (.14)	5.31	2.06 (.13)	2.82	2.12 (.14)	5.91	2.33 (.15)	16.36	2.65 (.20)	32.61
τ_{00} (2.0)	2.20 (.32)	10.20	2.18 (.34)	9.20	2.21 (.32)	10.36	2.20 (.34)	9.86	2.11 (.33)	5.40	2.02 (.37)	.78
τ_{10} (-0.2)	-.18 (.14)	-9.71	-.18 (.15)	-11.26	-.19 (.14)	-6.96	-.19 (.15)	-6.57	-.17 (.15)	-17.05	-.15 (.16)	-23.86
τ_{11} (1.0)	1.01 (.11)	.54	1.01 (.11)	.65	1.01 (.11)	1.06	1.01 (.11)	1.48	1.00 (.11)	.48	1.01 (.12)	.73
τ_{20} (0.0)	.00 (.07)		.00 (.06)		-.01 (.07)		-.01 (.06)		-.02 (.08)		-.05 (.08)	
τ_{21} (0.0)	.00 (.04)		.00 (.04)		.00 (.04)		.00 (.04)		.00 (.05)		.01 (.05)	
τ_{22} (0.10)	.08 (.03)	-17.37	.07 (.02)	-27.99	.08 (.03)	-18.30	.07 (.02)	-29.78	.10 (.04)	.62	.10 (.03)	3.98

Note. *R*=Replications per condition.

Table 8

Average Parameter Estimates for Model 2 Missing Data Conditions, 10 Repeated Measures

Parameter (Pop Value)	Listwise											
	MCAR				MAR				MNAR			
	20 (*R*=1000)		40 (*R*=999)		20 (*R*=1000)		40 (*R*=1000)		20 (*R*=1000)		40 (*R*=996)	
	Mean	Rel Bias	Mean	Rel Bias	Mean	Rel Bias	Mean	Rel Bias	Mean	Rel Bias	Mean	Rel Bias
Intercept, γ_{00}(5.0)	4.42 (.22)	-11.63	4.37 (.23)	-12.59	4.40 (.21)	-11.91	4.36 (.22)	-12.90	4.41 (.22)	-11.85	4.08 (.21)	-18.40
TIC, γ_{02} (1.0)	1.00 (.21)	.48	1.01 (.22)	.58	1.00 (.21)	.44	1.00 (.22)	.40	1.00 (.22)	.24	1.00 (.23)	-.33
Time$_{ti}$, γ_{10}(.5)	.49 (.11)	-1.38	.49 (.11)	-2.38	.49 (.11)	-1.76	.49 (.11)	-2.45	.49 (.11)	-1.02	.49 (.10)	-2.60
Time$_{ti}$*$\bar{z}_i$, γ_{11} (0.0)	.00 (.05)		.01 (.05)		.01 (.04)		.01 (.05)		-.01 (.06)		-.02 (.07)	
Time$_{ti}$*$\dot{z}_{ti}$, γ_{30}(.05)	.05 (.01)	.57	.05 (.01)	.37	.05 (.01)	.76	.05 (.01)	.56	.05 (.01)	.94	.05 (.01)	4.19
$\bar{z}_i$, γ_{01}(.2)	.29 (.08)	43.99	.31 (.08)	55.99	.31 (.08)	55.20	.35 (.08)	72.83	.14 (.10)	-30.82	.23 (.11)	12.81
$\dot{z}_{ti}$, γ_{20}(1.0)	1.00 (.05)	-.07	1.00 (.06)	-.07	1.00 (.05)	-.09	1.00 (.06)	-.01	1.00 (.06)	-.20	.99 (.07)	-.98
σ^2 (2.0)	2.00 (.08)	.13	2.00 (.09)	.13	2.01 (.08)	.26	2.01 (.10)	.33	2.00 (.08)	.23	2.01 (.09)	.27
τ_{00} (2.0)	2.14 (.27)	6.81	2.17 (.29)	8.72	2.14 (.27)	7.09	2.20 (.30)	9.97	2.19 (.28)	9.42	2.31 (.31)	15.41
τ_{10} (-0.2)	-.19 (.11)	-6.02	-.18 (.12)	-8.45	-.19 (.11)	-6.22	-.19 (.12)	-5.16	-.19 (.12)	-6.98	-.19 (.13)	-6.32
τ_{11} (1.0)	1.01 (.09)	1.47	1.01 (.09)	1.48	1.02 (.09)	1.51	1.02 (.10)	1.58	1.02 (.09)	1.65	1.02 (.10)	1.62
τ_{20} (0.0)	.00 (.06)		.00 (.07)		.00 (.06)		.00 (.07)		-.03 (.07)		-.05 (.08)	
τ_{21} (0.0)	.00 (.03)		.00 (.04)		.00 (.03)		.00 (.04)		.00 (.04)		.00 (.04)	
τ_{22} (0.10)	.10 (.02)	-2.14	.10 (.03)	-2.88	.10 (.02)	-3.05	.10 (.03)	-3.51	.10 (.03)	-2.33	.10 (.04)	-3.67
	Multiple Imputation											
	20 (*R*=1000)		40 (*R*=602)		20 (*R*=1000)		40 (*R*=996)		20 (*R*=1000)		40 (*R*=980)	
Intercept, γ_{00}(5.0)	4.45 (.22)	-11.02	4.45 (.22)	-11.08	4.46 (.22)	-10.82	4.46 (.22)	-10.84	4.45 (.22)	-10.91	4.46 (.21)	-10.72
TIC, γ_{02} (1.0)	1.01 (.21)	.58	1.01 (.21)	.74	1.00 (.21)	.40	1.01 (.22)	.57	1.00 (.21)	.42	1.00 (.22)	.46
Time$_{ti}$, γ_{10}(.5)	.50 (.11)	-.89	.50 (.11)	-.91	.49 (.11)	-1.68	.49 (.11)	-1.59	.49 (.11)	-2.23	.50 (.11)	-.21
Time$_{ti}$*$\bar{z}_i$, γ_{11} (0.0)	.00 (.05)		.00 (.05)		.01 (.05)		.02 (.05)		.02 (.05)		.03 (.06)	
Time$_{ti}$*$\dot{z}_{ti}$, γ_{30}(.05)	.05 (.01)	-1.69	.05 (.01)	-6.57	.05 (.01)	-.99	.05 (.01)	-2.71	.06 (.01)	27.47	.08 (.01)	68.72
$\bar{z}_i$, γ_{01}(.2)	.27 (.08)	36.21	.27 (.08)	36.83	.26 (.08)	30.98	.24 (.09)	22.43	.28 (.09)	39.77	.28 (.11)	41.54
$\dot{z}_{ti}$, γ_{20}(1.0)	.99 (.05)	-1.24	.97 (.06)	-3.16	.98 (.05)	-2.05	.94 (.06)	-5.54	1.01 (.06)	.84	.92 (.07)	-7.54
σ^2 (2.0)	2.08 (.07)	4.02	2.19 (.09)	9.32	2.09 (.08)	4.45	2.20 (.09)	10.12	2.31 (.09)	15.54	2.64 (.13)	32.25
τ_{00} (2.0)	2.10 (.26)	4.92	2.07 (.27)	3.63	2.10 (.26)	5.02	2.09 (.28)	4.73	2.06 (.27)	3.24	2.02 (.29)	.78
τ_{10} (-0.2)	-.19 (.11)	-5.61	-.19 (.11)	-6.21	-.19 (.11)	-4.40	-.20 (.12)	-2.37	-.19 (.11)	-5.17	-.19 (.12)	-6.72
τ_{11} (1.0)	1.01 (.09)	1.44	1.01 (.09)	1.11	1.02 (.09)	1.55	1.02 (.09)	1.71	1.02 (.09)	1.56	1.02 (.09)	1.59
τ_{20} (0.0)	.00 (.05)		.00 (.04)		.00 (.05)		-.01 (.04)		-.01 (.06)		-.02 (.06)	
τ_{21} (0.0)	.00 (.03)		.00 (.03)		.00 (.03)		.00 (.03)		.00 (.03)		.00 (.03)	
τ_{22} (0.10)	.08 (.02)	-23.25	.06 (.02)	-40.05	.08 (.02)	-23.80	.06 (.02)	-41.11	.09 (.02)	-10.59	.08 (.02)	-20.37

Note. *R*=Replications per condition.

Table 9

Model 2 RMSE

	Listwise													
	5 Repeated Measures							10 Repeated Measures						
	Complete	MCAR		MAR		MNAR		Complete	MCAR		MAR		MNAR	
Parameter (Pop Value)		20	40	20	40	20	40		20	40	20	40	20	40
Intercept, γ_{00}(5.0)	.70	.75	.82	.75	.79	.80	1.04	.59	.62	.67	.63	.68	.63	.94
TIC, γ_{02} (1.0)	.22	.23	.23	.23	.23	.22	.23	.21	.21	.22	.21	.22	.22	.23
Time_{ti} , γ_{10}(.5)	.12	.12	.12	.12	.12	.12	.12	.11	.11	.11	.11	.11	.11	.10
Time_{ti}*$\bar{z}_{\text{i}}$, γ_{11} (0.0)	.05	.05	.05	.05	.06	.06	.07	.05	.05	.05	.05	.05	.06	.07
Time_{ti}*$\dot{z}_{ti}$, γ_{30}(.05)	.03	.03	.04	.03	.04	.04	.04	.01	.01	.01	.01	.01	.01	.01
$\bar{z}_{\text{i}}$, γ_{01}(.2)	.15	.18	.21	.20	.23	.12	.19	.11	.12	.14	.14	.17	.12	.11
$\dot{z}_{ti}$, γ_{20}(1.0)	.07	.08	.09	.08	.09	.08	.09	.05	.05	.06	.05	.06	.06	.07
σ^2 (2.0)	.12	.14	.17	.14	.16	.14	.16	.07	.08	.09	.08	.10	.08	.09
τ_{00} (2.0)	.38	.42	.50	.43	.51	.46	.58	.27	.30	.34	.31	.36	.34	.44
τ_{10} (-0.2)	.14	.15	.16	.15	.17	.15	.18	.11	.11	.12	.12	.12	.12	.13
τ_{11} (1.0)	.11	.11	.12	.11	.13	.11	.13	.09	.09	.09	.10	.10	.09	.10
τ_{20} (0.0)	.07	.09	.10	.09	.10	.10	.13	.05	.06	.07	.06	.07	.07	.10
τ_{21} (0.0)	.05	.05	.06	.05	.06	.06	.07	.03	.03	.04	.03	.04	.04	.04
τ_{22} (0.10)	.04	.05	.05	.04	.05	.05	.06	.02	.02	.03	.02	.03	.03	.04
	Multiple Imputation													
		MCAR		MAR		MNAR			MCAR		MAR		MNAR	
Intercept, γ_{00}(5.0)	.70	.70	.71	.69	.70	.70	.74	.59	.59	.60	.58	.58	.59	.58
TIC, γ_{02} (1.0)	.22	.22	.23	.22	.23	.22	.22	.21	.21	.21	.21	.22	.21	.22
Time_{ti} , γ_{10}(.5)	.12	.12	.12	.12	.12	.12	.12	.11	.11	.11	.11	.11	.11	.11
Time_{ti}*$\bar{z}_{\text{i}}$, γ_{11} (0.0)	.05	.05	.05	.05	.07	.08	.12	.05	.05	.05	.05	.05	.06	.07
Time_{ti}*$\dot{z}_{ti}$, γ_{30}(.05)	.03	.03	.04	.03	.04	.05	.10	.01	.01	.01	.01	.01	.02	.04
$\bar{z}_{\text{i}}$, γ_{01}(.2)	.15	.15	.15	.14	.13	.14	.15	.11	.11	.11	.10	.10	.12	.14
$\dot{z}_{ti}$, γ_{20}(1.0)	.07	.08	.09	.08	.10	.09	.16	.05	.05	.07	.06	.08	.06	.10
σ^2 (2.0)	.12	.14	.18	.14	.18	.36	.68	.07	.11	.21	.12	.22	.32	.66
τ_{00} (2.0)	.38	.38	.39	.38	.39	.34	.37	.27	.28	.28	.28	.29	.27	.29
τ_{10} (-0.2)	.14	.15	.15	.15	.15	.15	.17	.11	.11	.11	.11	.12	.11	.12
τ_{11} (1.0)	.11	.11	.11	.11	.11	.11	.12	.09	.09	.09	.09	.10	.09	.10
τ_{20} (0.0)	.07	.07	.06	.07	.06	.08	.10	.05	.05	.04	.05	.04	.06	.06
τ_{21} (0.0)	.05	.04	.04	.04	.04	.05	.05	.03	.03	.03	.03	.03	.03	.03
τ_{22} (0.10)	.04	.04	.04	.03	.04	.04	.03	.02	.03	.04	.03	.04	.02	.03

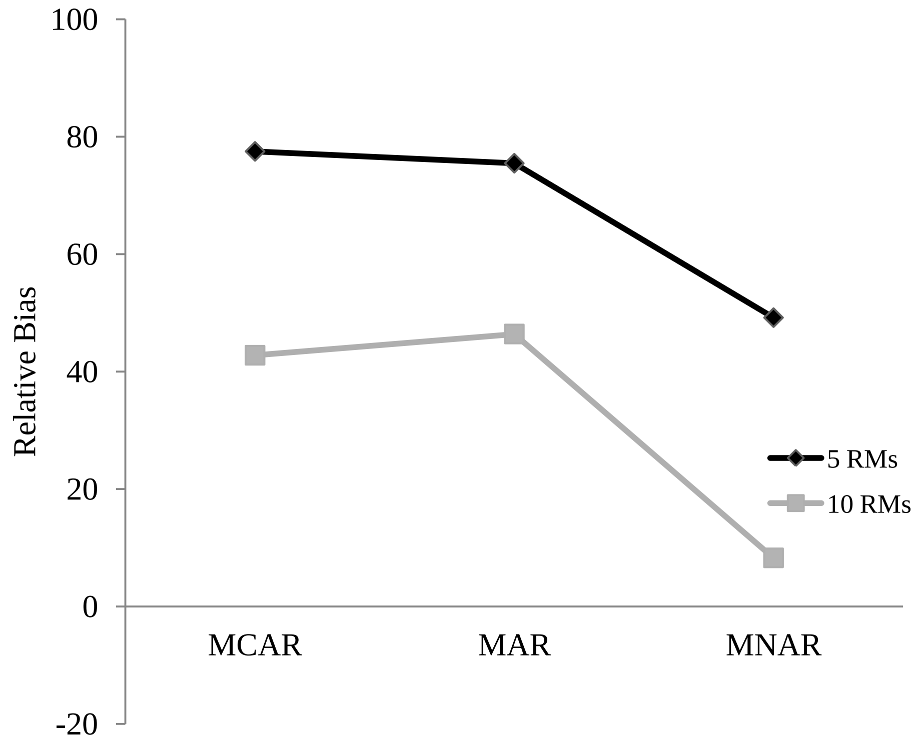

Figure 1. Effects of missingness mechanism and number of repeated measures on relative bias in the between-person effect (γ_{01}) in Model 1.

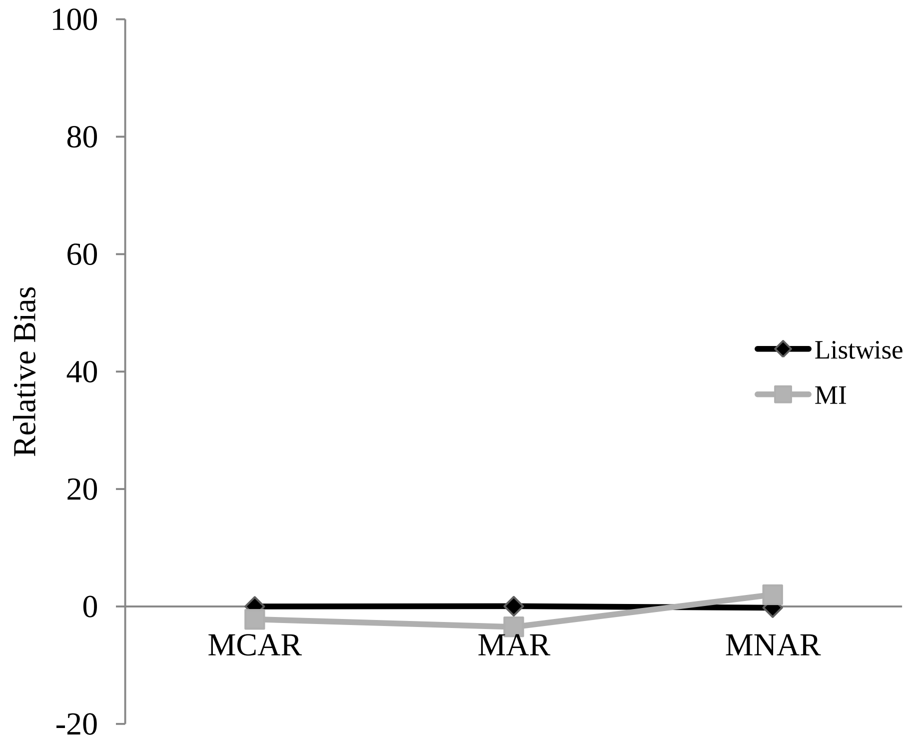

Figure 2. Effects of mechanism and MI on relative bias in the within-person effect (γ_{20}) in Model 1.

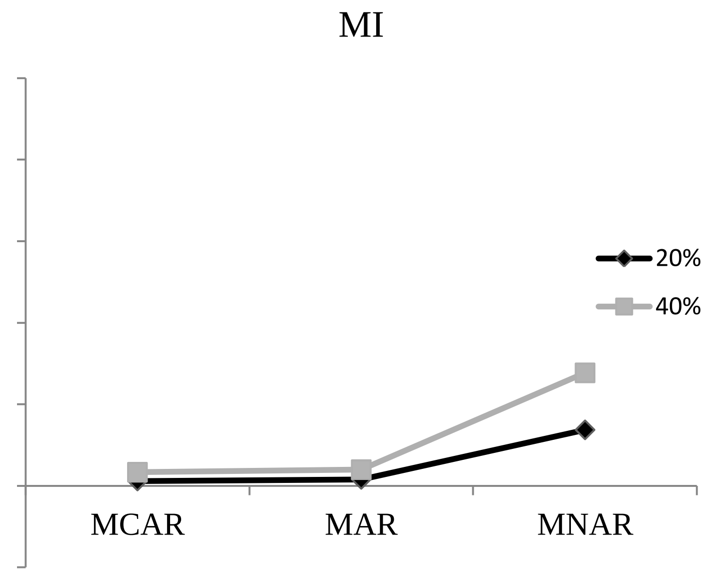

Figure 3. Effects of proportion missing and mechanism on relative bias in the residual variance estimate (σ^2) after MI in Model 1.

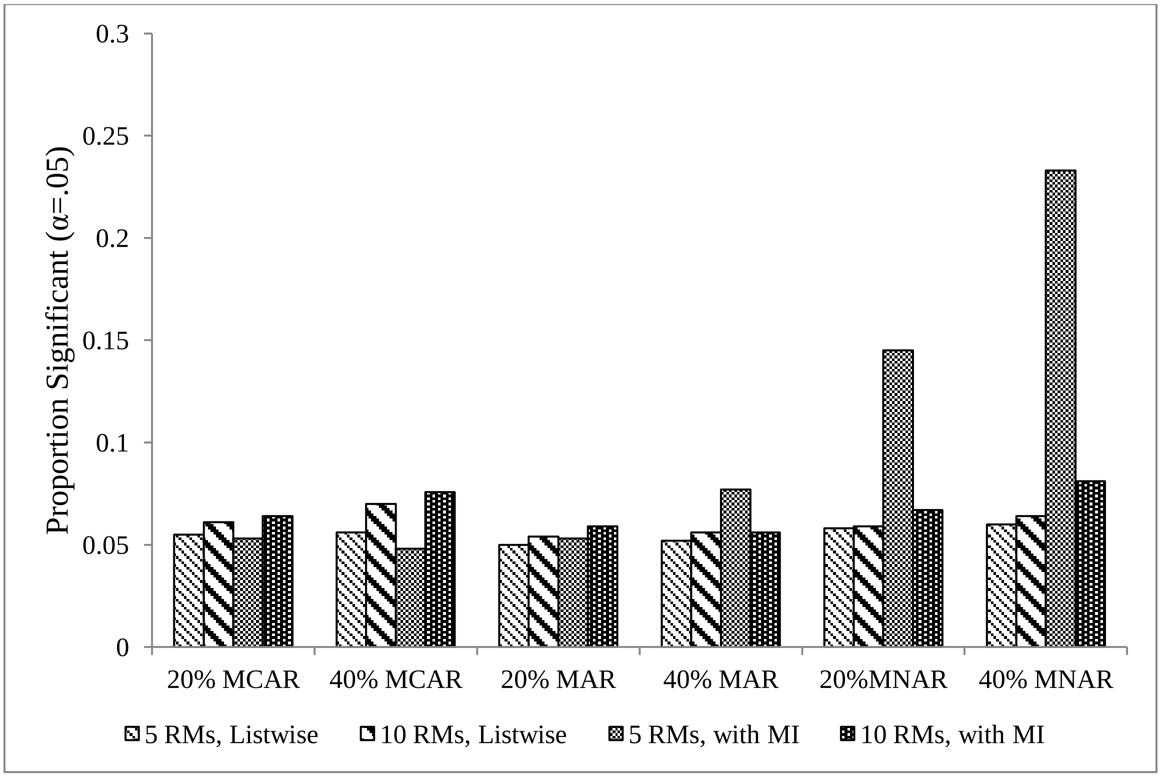

Figure 4. Proportion of significant estimates of the interaction between time and the between-person effect (γ_{11}) in Model 1.

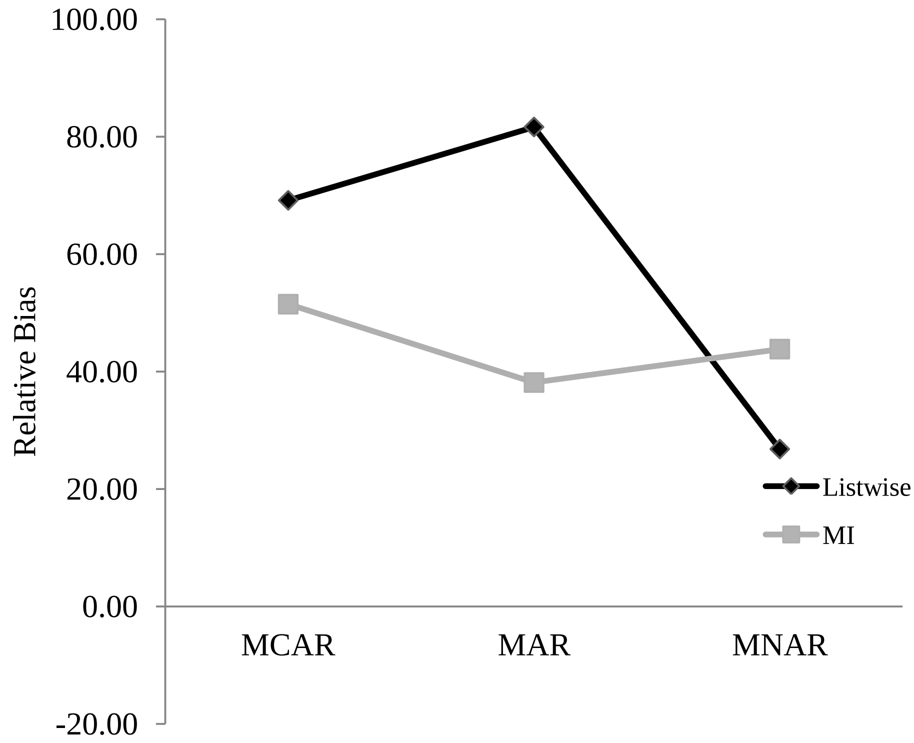

Figure 5. Effects of missingness mechanism and MI on relative bias in the between-person effect (γ_{01}) in Model 2.

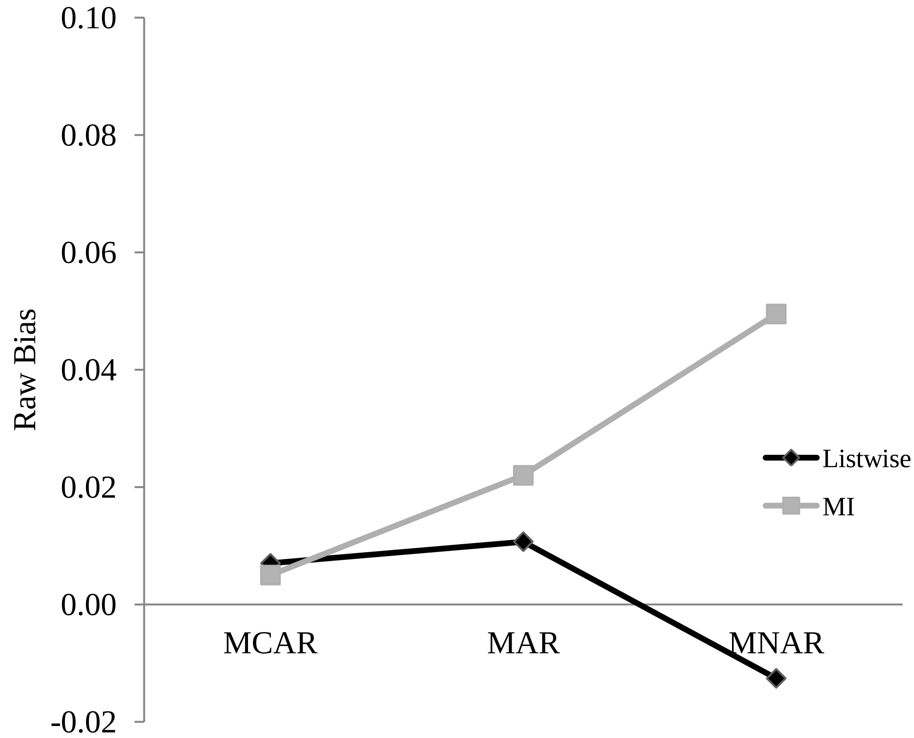

Figure 6. Effects of missingness mechanism and MI on raw bias in the interaction between time and the between-person effect (γ_{11}) in Model 2.

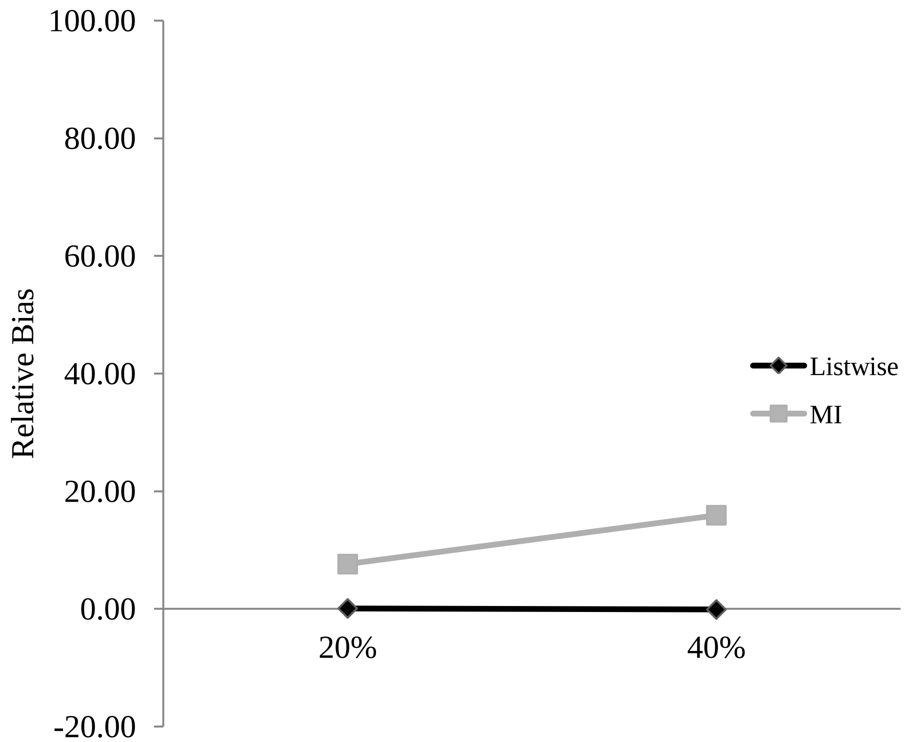

Figure 7. Effects of percent missing and MI on relative bias in the residual variance estimate (σ^2) in Model 2.

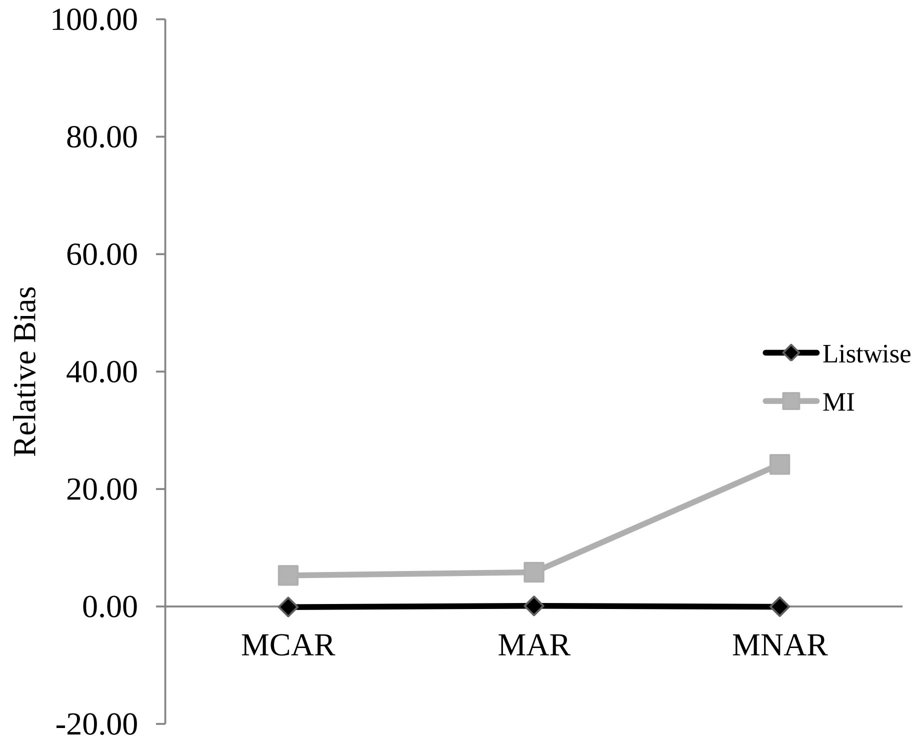

Figure 8. Effects of missingness mechanism and MI on relative bias in the residual variance estimate (σ^2) in Model 2.

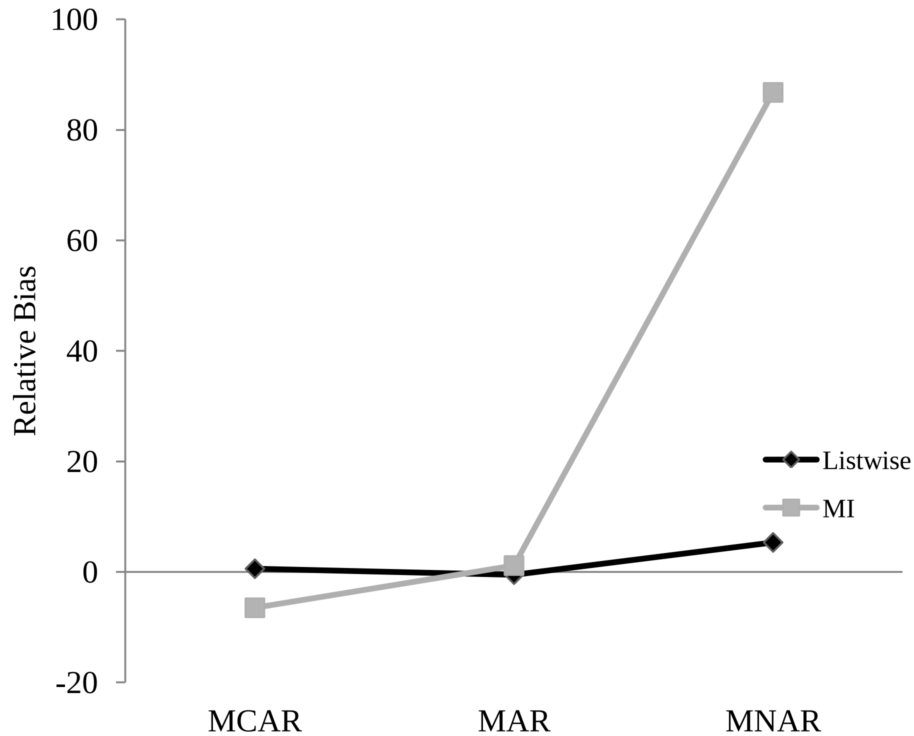

Figure 9. Effects of missingness mechanism and MI on relative bias in the interaction between time and the within-person effect (γ_{30}) in Model 2.

REFERENCES

Allison, P. D. (2003). Missing data techniques for structural equation modeling. *Journal of Abnormal Psychology, 112*(4), 545-557. doi:10.1037/0021-843X.112.4.545

Armeli, S., Todd, M., & Mohr, C. (2005). A daily process approach to individual differences in stress-related alcohol use. *Journal of Personality, 73,* 1-30.

Belin T.R, Liu M, & Taylor J.M.G. (2000). Multiple imputation and posterior simulation for multivariate missing data in longitudinal studies. *Biometrics, 56*(4), 1157-1157. doi:10.1111/j.0006-341X.2000.01157.x

Casella, G., & George, E. I. (1992). Explaining the Gibbs Sampler. *The American Statistician, 46*, 167-174.

Collins, L. M., Schafer, J. L., & Kam, C. M. (2001). A comparison of inclusive and restrictive strategies in modern missing data procedures. *Psychological Methods, 6*(4), 330-351. doi:10.1037/1082-989X.6.4.330

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, N.J.: L. Erlbaum Associates.

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

Curran, P. J., & Bauer, D. J. (2011). The disaggregation of within-person and between-person effects in longitudinal models of change. *Annual Review of Psychology, 62,* 583.

Curran, P.J., Lee, T.H., MacCallum, R.A., & Lane, S. (in press). Disaggregating within-person and between-person effects in multilevel and structural equation growth models. To appear in *Advances in Longitudinal Methods in the Social and Behavioral Sciences* (G. Hancock, Ed.) Charlotte, NC: Information Age.

Enders, C.K. (2010). *Applied Missing Data Analysis*. New York: Guilford Press.

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*(3), 430-57. doi:10.1207/S15328007SEM0803_5

Enders, C., & Gottschall, A. (2011). Multiple imputation strategies for multiple group structural equation models. *Structural Equation Modeling-A Multidisciplinary Journal, 18*(1), 35-54. doi:10.1080/10705511.2011.532695

Fitzmaurice, G.M., Laird, N.M., & Ware, J.H. (2011). *Applied Longitudinal Analysis*. Hoboken, N.J. : Wiley.

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*(3), 206-213. doi:10.1007/s11121-007-0070-9

Hardy, J. K., Crofford, L. J., & Segerstrom, S. C. (2011). Goal conflict, distress, and pain in women with fibromyalgia: A daily diary study. *Journal of Psychosomatic Research*, 70(6), 534-540.

Hoffman, L., & Stawski, R. (2009). Persons as contexts: Evaluating between-person and within-person effects in longitudinal analysis. *Research in Human Development, 6*(2-3), 97-120. doi:10.1080/15427600902911189

Hussong, A. M., Cai, L., Curran, P. J., Flora, D. B., Chassin, L. A., & Zucker, R. A. (2008). Disaggregating the distal, proximal, and time-varying effects of parent alcoholism on children's internalizing symptoms. *Journal of Abnormal Child Psychology, 36*(3), 335-346. doi:10.1007/s10802-007-9181-9

Jahng, S., Solhan, M. B., Tomko, R. L., Wood, P. K., Piasecki, T. M., & Trull, T. J. (2011). Affect and alcohol use: An ecological momentary assessment study of outpatients with borderline personality disorder. *Journal of Abnormal Psychology, 120*(3), 572-584.

Kreft, I., Deleeuw, J., & Aiken, L. (1995). The effect of different forms of centering in hierarchical linear-models. *Multivariate Behavioral Research, 30*(1), 1-21.

Lüdtke, O., Marsh, H. W., Robitzsch, A., Trautwein, U., Asparouhov, T., & Muthén, B. (2008). The multilevel latent covariate model: A new, more reliable approach to group-level effects in contextual studies: A new, more reliable approach to group-level effects in contextual studies. *Psychological Methods, 13*(3), 203-229. doi:10.1037/a0012869

Marsh, H., Lüdtke, O., Robitzsch, A., Trautwein, U., Asparouhov, T., Muthen, B., & Nagengast, B. (2009). Doubly-latent models of school contextual effects: Integrating multilevel and structural equation approaches to control measurement and sampling error. *Multivariate Behavioral Research*, *44*(6), 764-802. doi:10.1080/00273170903333665

Newman, D. A. (2003). Longitudinal modeling with randomly and systematically missing data: A simulation of ad hoc, maximum likelihood, and multiple imputation techniques. *Organizational Research Methods, 6*(3), 328-362. doi:10.1177/1094428103254673

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*(1), 53-79. doi:10.1016/S0377-2217(02)00578-7

Preacher, K., Zhang, Z., & Zyphur, M. (2011). Alternative methods for assessing mediation in multilevel data: The advantages of multilevel SEM. *Structural Equation Modeling-A Multidisciplinary Journal*, *18*(2), 161-182. doi:10.1080/10705511.2011.557329

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

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

Rubin, D.B. (1987) *Multiple Imputation for Nonresponse in Surveys*. J. Wiley & Sons, New York.

Schafer, J. L. (1997). *Analysis of Incomplete Multivariate Data*. London: Chapman & Hall.

Schafer, J.L. (2001) Multiple imputation with PAN. In A.G. Sayer and L.M. Collins (Eds.), *New Methods for the Analysis of Change*, Washington, DC: American Psychological Association, 355-377.

Schafer, J. L. (2003). Multiple imputation in multivariate problems when the imputation and analysis models differ. *Statistica Neerlandica*, *57*(1), 19-35. doi:10.1111/1467-9574.00218

Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods*, *7*(2), 147-177. doi:10.1037/1082-989X.7.2.147

Shin, Y., & Raudenbush, S. W. (2007). Just-identified versus overidentified two-level hierarchical linear models with missing data. *Biometrics, 63*(4), 1262-1262. doi:10.1111/j.1541-0420.2007.00818.x

Shin, Y. & Raudenbush, S. W. (2010). A latent cluster-mean approach to the contextual effects model with missing data. *Journal of Educational and Behavioral Statistics, 35*(1), 26-53. doi:10.3102/1076998609345252

Shin, Y. & Raudenbush, S.W. (Submitted). Efficient analysis of q-level nested hierarchical general linear models given ignorable missing data.

Shrout, P. E., & Lane, S. P. (2012). Reliability. *APA Handbooks in Psychology: APA Handbook of Research Methods in Psychology* (Volume 1). Washington, DC: American Psychological Association.