Comparing Different Planned Missingness Designs in Longitudinal Studies

Josephine Wood · **Gregory J. Matthews** · **Nicholas Illenberger** · **Jennifer Pellowski** · **Ofer Harel**

Abstract Planned Missingness (PM) designs, in which researchers deliberately collect only partial data, have enjoyed a recent growth in popularity. Among other benefits these designs have been proven capable of reducing the study costs and alleviating participant burden. Past research has shown that Split Form PM designs can be effective in simplifying complex surveys while Wave Missingness PM designs act similarly for Longitudinal studies. However, less work has been done to inform how to implement PM structures into studies which incorporate elements of both survey and longitudinal designs. Specifically, in studies where a questionnaire is given to participants at multiple measurement occasions the best way to design missingness is still unclear. To address this deficiency, data in this hybrid format was simulated under both Split Form and Wave Missingness PM structures. Multiple Imputation techniques were applied to estimate a multilevel logistic model in each of the simulations. Estimated parameters were compared to the true values to see which PM design allowed us to best capture the true model. The results of this study indicate that, compared to the Split Form Design, the Wave Missingness design consistently performed less effectively in capturing the multilevel model. Thus, in the context of longitudinal surveys this study recommends the use of Split Form missingness designs, which performs well under a number of different conditions.

J. Wood · G. J. Matthews
Loyola University Chicago
1032 W. Sheridan Rd.
Chicago IL, 60660 USA
Tel: (773)508-3558

N. Illenberger · O. Harel
University of Connecticut
215 Glennbrook Rd. Unit 420
Storrs CT, 06269-4120 USA Tel: (860)486-6989
E-mail: ofer.harel@uconn.edu

J. Pellowski
International Health Institute
Brown School of Public Health

J. Pellowski
Department of Psychiatry and Human Behavior
The Warren Alpert Medical School of Brown University.

1 Introduction

Planned Missingness (PM) designs are a type of experimental design in which some of the data is predetermined to be missing. While it may seem counterproductive to deliberately omit data from the collection process, PM designs have been shown to reduce both participant burden [Graham et al., 2006, Saris et al., 2004] and the costs of a study [Graham et al., 2001]. Moreover, PM designs can actually increase the quality of collected data by increasing the breadth of a study [Little and Rhemtulla, 2013] as well as decreasing the effects of assessment reactivity and rates of unplanned missingness Harel et al. [2015]. We further note that while PM designs decrease the efficiency of estimators [Rhemtulla et al., 2016], decreasing the cost per participant may allow researchers to enroll more participants in a study, offsetting some of the efficiency loss.

There have been many different incarnations of the PM design, each crafted to best suit the study at hand. For example, PM designs have been used in the study of longitudinal growth curves [Graham et al., 2001, McArdle and Woodcock, 1997], business research [Shoemaker, 1973], behavioral studies [Harel et al., 2015], biostatistics [Andrés Houseman and Milton, 2006, Wacholder et al., 1994], and educational assessments [Sirotnik and Wellington, 1977, Zeger and Thomas, 1997]. This paper focuses on two of the simplest and most applicable PM designs; the Split Form design [Raghunathan and Grizzle, 1995] and the Wave Missingness design [Little and Rhemtulla, 2013]. While this paper covers these designs in more detail later, broadly speaking, the Split Form design entails measuring fewer points on a single measurement occasion [Raghunathan and Grizzle, 1995]. The Wave Missingness design, on the other hand, requires missing an entire measurement occasion within a longitudinal study [Graham et al., 2001].

The issue we concern ourselves with arises when a longitudinal study requires multiple items to be collected at each measurement occasion. When implementing a multi-level regression model in this situation, we investigate which planned missingess design is more efficient and accurate. Previous research indicates that for the estimation of correlated latent growth curve models use of the three-form PM design, a special case of Split Form designs [Raghunathan and Grizzle, 1995], resulted in low bias on parameter estimates or standard error and high efficiency when compared with wave missingness designs [Rhemtulla et al., 2014]. As our model of interest is a mixed effects model, however, we can not be sure that this result will hold.

To answer our question we performed a simulation study from data obtained in longitudinal research on HIV medication adherence and alcohol-interactivity beliefs [Pellowski et al., 2016]. For the PM designs mentioned we simulated sample data and then used multiple imputation (MI) [Little and Rubin, 2014] techniques to analyze the data. The PM methods were subsequently evaluated through the assessmemt of the average bias, percent bias, mean squared error, coverage rates, confidence interval length, and estimated fraction of missing information (FMI) [Little and Rubin, 2014] for parameter estimates in a multilevel model. To make the results generalizable, the experiment was repeated under multiple scenarios: increased sample size, low inter-survey correlation, and high time effect.

To delineate the remainder of this paper, in Section 2 we go into more depth on the two PM designs previously mentioned and then move into a discussion on the modern Multiple Imputation technique for dealing with missing data. Next, in Section 3 we briefly describe

the dataset we worked with and our method of simulation before moving on to the results of our analyses in Section 4 and a discussion in the final section.

2 Designs and Methods

2.1 Split Form Design

Split Form missingness designs [Raghunathan and Grizzle, 1995], also known as Multi-Form [Little and Rhemtulla, 2013] or Matrix Sampling designs [Thomas et al., 2006], are very aptly named. Instead of giving participants one large survey, researchers create multiple forms, each containing a subset of the full survey's questions. How these subsets are chosen varies. A common form setup is three-form Block Design [Graham et al., 1996] in which the full survey is split into four blocks: X, A, B, and C. Each form of the survey includes the X block as well as two of the remaining blocks. A participant is then given a different form each day of the study. Designing the forms in this manner allows researchers to estimate the relationship between any chosen items, regardless of their blocks. For example, this allows the relationship between an item in block A and another in block C to be estimated and examined.

An important consideration here is question distribution. The X-block, being included in every form, should contain either the variables of greatest interest or those with the most predictive ability. Placing central items in the X-block ensures they are always collected and minimizes the efficiency loss of our analyses [Thomas et al., 2006]. On the other hand, placing the most predictive variables in the X-block allows Multiple Imputation to more effectively impute missing values from the other blocks [Gottschall et al., 2012]. Similarly, it is good practice to place variables which are predictive of each other in separate blocks [Gottschall et al., 2012]. While this approach provides researchers more freedom when forming the surveys, the puzzle of question distribution introduces new concerns. When the question groupings are not obvious, pilot studies may be necessary to determine the predictive potential of and relationships between variables. The cost and burden associated with an item must also be considered when deciding which block to place it in. For example, if a critical item is overly expensive or difficult to obtain we may be better off excluding it from the X-block [Little and Rhemtulla, 2013].

In addition, the number of blocks must be determined. Including more blocks in the design allows further increased missingness [Graham et al., 2006]. While increasing missingness may allow for further reductions in burden and cost, researchers must consider the disadvantages of more blocks. As more blocks are added, the number of possible forms increases. For example, consider two research teams. The first team divides their survey into an X-block and three other blocks of equal size, each containing 25% of the survey. The second uses an X-block and nine others with 10% of the items each. If the teams decide that each form will contain the X-block and two of the other blocks, then the second team will have 36 **(28? 8 choose 2?)** total forms while the first team can proceed with 3. **This is all assuming that the blocks represent the optimal question grouping. If this is not known, to account for possible relationships and correlation between questions, the teams would need to consider all possible blocks and create forms for every block configuration. For a 7 question survey, with any one question X block and the other 6 split into three blocks, there are 75 possible block configurations. For each configuration, there are 3 possible forms. So, a team would need to consider 225 different forms.**

Instead of creating 225 distinct forms, a research team can follow a alternative Split Form method.

This method, hereafter referred to as Split Form planned missingness and illustrated in 1, involves meaningfully selecting an X block of n questions and then randomly excluding k questions on each day. In a way treating each item as its own block [Silvia et al., 2014]. This is simple in theory, but for traditional paper surveys this is very difficult to execute. Fortunately, recent technological advances have made the use of electronic questionaires more readily accessible. In this environment form randomization can be automated and the creation of numerous unique forms is no longer an issue. Using this method, researchers reap the benefits of including all possible variable relationships through increased forms without the difficulty caused by manually building these forms.

2.2 Wave Missingness Design

While Split Form missingness can be applied to many experimental designs, the Wave Missingness design [Little and Rhemtulla, 2013] is unique to studies with multiple measurement occasions, or "waves". Unlike the Split Form design, where participants miss a subset of items in a single measurement occasion, the Wave Design entails randomizing participants to miss a measurement occasion entirely. An example of a Wave Missing design is shown in Figure 1 where data is collected monthly over the course of 5 months. However, each participant is only measured at three of these occasions.

A relative advantage of the Wave Missingness design over the Split form design is increased control over the level of missingness. For studies with many measurement occasions, it is easy to make small changes to the amount of missing data. For example, the data described by Pellowski et al. [2016] records measurements for 45 days. Missing eleven days here would result in 24.4% missingness while missing twelve gives us 26.7%. Further, Wave Missingness designs eclipse Split Form designs in terms of cost reduction [Little and Rhemtulla, 2013].

Another advantage of this design is many longitudinal study designs already use the Wave Missingness structure [Little and Rhemtulla, 2013]. Cohort Sequential Designs require cohorts of different ages to be followed longitudinally in such a way that the information from one cohort can be used to impute values for another [Duncan et al., 2013, Little, 2013]. For example, there were three cohorts of ages 10, 11, and 12 at baseline and measured annually for four years. Although measurement only occurs at four waves we can frame this as a six-wave study where the 10-year old cohort has missing values at waves five and six, the 11-year old cohort at waves one and six, and the 12-year old cohort at waves one and two. Because the 10-year old cohort is the only group observed at the first wave, the missing values for the other groups are imputed using this information. By continuing in this manner the data that was collected over a four year period is extended it to cover six years, accelerating the study. The setup of this type of study is demonstrated in Table 1, where "1" and "0" respectively indicate an included and excluded wave.

Table 1: Cohort Sequential Design

Participant	Age					
	10	11	12	13	14	15
1	1	1	1	1	0	0
2	0	1	1	1	1	0
3	0	0	1	1	1	1

Developmental time-lag models [McArdle and Woodcock, 1997] use the Wave Missingness structure, as well. As shown in Table 2, researchers observe participants at only two measurement occasions, varying the length between occasions for each participant, and then treat this two-time-point data as multi-time-point data with missing values. Under this framework they impute these missing values and estimate growth curve models. Estimating this type of model would normally require measuring participants at many time points. However, under the Wave Missingness design this is accomplished with only two measurement occasions per participant. These designs exemplify how useful the Wave Missingness structure is in accelerating longitudinal studies.

Table 2: Developmental Time-Lag Setup

Participant	Wave			
	1	2	3	4
1	1	1	0	0
2	1	0	1	0
3	1	0	0	1

2.3 Altered Split Form

The Split Form planned missingness method can be modified to include the benefits of complete cases seen in Wave Missingness in what may be called Altered Split Form. In this method, the researcher asks participants all questions on the first n days of the study. Thereafter, the survey design reverts to regular Split Form. This method allows the researcher to collect n complete cases from each participant, aiding in the imputation process after surveys are administered. Altered Split Form, and the other planned missingness methods, is illustrated in Figure 1 below.
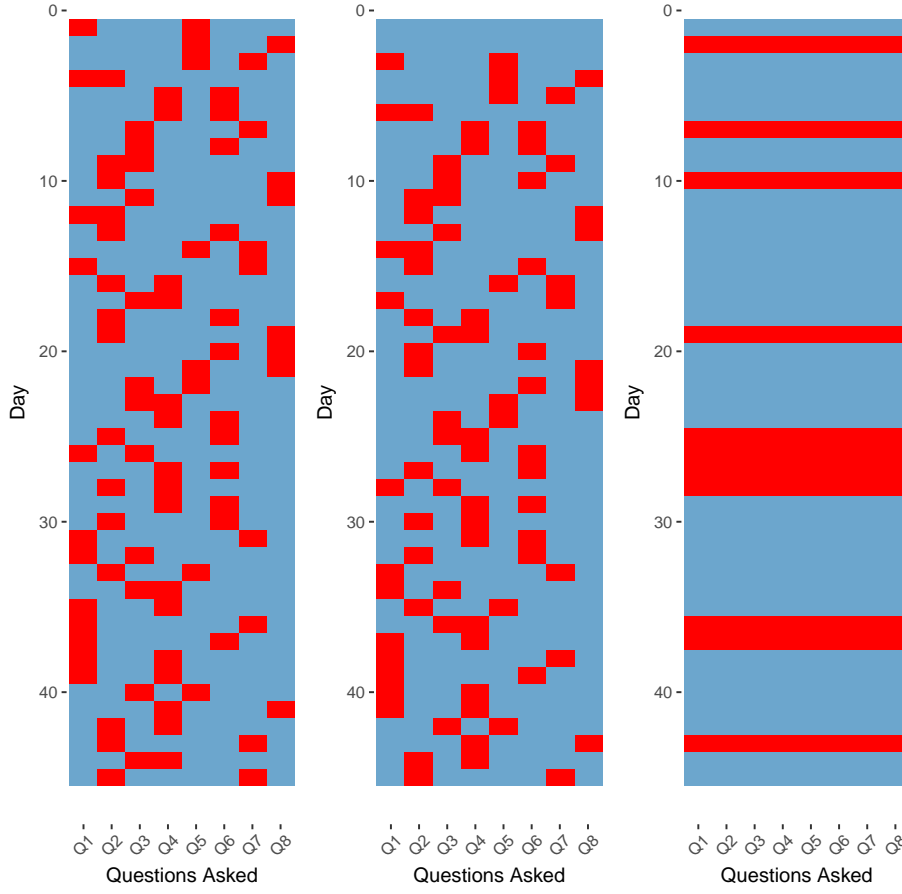
Fig. 1: Low missingness

2.4 Multiple Imputation

2.4.1 Missing Data Mechanisms

Before describing Multiple Imputation and how it can be applied to PM designs it is important to first understand how data become incomplete. Suppose we have a dataset, Y_{com}, with some missing values. We define the missingness, R, to be an indicator variable which equals 0 where Y_{com} is observed and 1 where the corresponding value in Y_{com} is missing. Missing data mechanisms describe the proposed probabilistic relationship between R and the values in Y_{com}, observed or unobserved [Little and Rubin, 2014].

The first missing data mechanism, missing at random (MAR), was defined by Rubin [1976]. Data are considered MAR when the missingness does not depend on unobserved values but may depend upon other observed variables. For instance, if men are less likely to

report their incomes than women but the level of income itself does not affect the probability of reporting, then we can say the missingness govern income is MAR.

Later, Little and Rubin [2014] described two additional mechanisms which lead to missing values: missing completely at random (MCAR) and missing not at random (MNAR) [Little and Rubin, 2014]. The first, MCAR, is used to describe a situation in which the missingness is independent of the observed or unobserved data. That is, the missing values are simply a random subset of the full data. On the other hand, data is said to be MNAR when the missingness is related to the unobserved data; in a survey people may be less willing to provide their income if it is extremely high or low, for example.

If we have prior knowledge concerning the relationship between unobserved values, indicated by R, and the observed values, then Multiple Imputation can yield efficient estimates even under MNAR [Harel and Zhou, 2007]. However, because this is not typically the case, MI is most often implemented when data is MCAR or MAR [Little and Rubin, 2014].

As a consequence of allowing researchers to decide which values in a dataset are observed and unobserved, PM designs grant control over the missing data mechanism. Because of this, MI is an appealing option for researchers utilizing a PM design. If implemented properly, data should conform to either the MCAR or MAR assumption. For instance, if medical researchers decide not to collect blood pressure on a completely random subset of patients there is no reason to believe the missing blood pressures differ significantly from the observed ones. This fact allows for the use of MI and other modern missing data techniques such as the Full Information Maximum Likelihood approach [Dempster et al., 1977]

2.4.2 The Multiple Imputation Procedure

Multiple Imputation is used to obtain statistically valid inferences in spite of missing data. MI works by replacing unobserved values with a set of plausible values in a way which allows us to account for uncertainty in the imputation process. Repeatedly using an imputation technique, every missing value in the incomplete dataset is replaced with a fixed number of imputed values. This process allows for the creation of multiple complete datasets, on which users can run complete-data procedures. If, for instance, we imputed M possible values for every missing value we would obtain M different completed datasets. Once created, the analysis of interest can be performed on each of them and, using rules defined by Little and Rubin [2014] and Rubin [2004], combine the results to obtain a single inference.

Suppose we are interested in estimating a certain parameter such as the mean of a variable or a regression coefficient. If θ is the parameter of interest, let $\widehat{\theta_1}, \widehat{\theta_2}, ... \widehat{\theta_M}$ be the estimates of θ and let $S_1, S_2, ... S_M$ be the corresponding estimated standard errors from the M imputed datasets. We can aggregate these results using the combining rules specified by Little and Rubin [2014] and Rubin [2004].

The overall MI estimate of θ is defined as,

$$\widehat{\theta} = \sum_{n=1}^{M} \widehat{\theta}_n. \tag{1}$$

From here we can compute the sample variance of this estimate by calculating the between-imputation component, $B_M = \frac{1}{(M-1)} \sum_{n=1}^{M} (\widehat{\theta}_n - \widehat{\theta})^2$, and the within-imputation component, $W_M = \frac{1}{(M)} \sum_{n=1}^{M} S_n^2$. These components are combined to obtain the overall sample variance of our MI estimate as follows:

$$T_M = W_M + \frac{M+1}{(M)} B_M.$$ (2)

These equations form the basis of Rubin's combining rules and provide unbiased parameter estimates accounting for the inherent uncertainty in the imputation process [Raghunathan, 2015]. For a more in depth explanation of these methods we refer readers to Little and Rubin [2014].

When using Multiple Imputation, researchers must decide on the number of imputations, M, to use. Generally, as the level of missingness decreases the number of imputations needed does as well. If less than 20% of the requisite data is missing, for instance, five or tien imputations will suffice [Raghunathan, 2015]. However, at its roots, MI is a simulation technique and thus, the greater M is the more stable our estimates become [Harel, 2007]. Graham et al. [2007] recommend that if computational power is not a limiting factor, then setting $M = 100$ is good practice.

3 Data and Simulations

3.1 Data

The dataset the simulations are based on originated in a study by Pellowski et al. [2016]. The data was collected in an attempt to determine the day-level relationship between alcohol use, its interactive toxicity beliefs, and adherence to antiretroviral therapy (ART) amongst patients with HIV. Previous studies note that alcohol use combined with the perception that mixing alcohol and ARTs can cause adverse effects, correlated with non-adherence to ARTs [Kalichman et al., 2009, 2012]. However, Pellowski et al. [2016] pointed out that past research collected data over extended periods of time, preventing the identification a day-level interaction. Unlike others, this study was able to determine whether alcohol use and missed medication doses occurred on the same day.

The study employed an observational cohort design, following sixty participants over forty-five consecutive days via a daily questionnaire. Questions in the survey concerned livelihood insecurities from the previous day as well as alcohol use. The analysis included eight particular questions: two binary questions on housing insecurity, four binary questions on food insecurity, and two questions concerning alcohol use (one binary variable and one count variable). The exact questions can be found in Table 3.

Table 3: Survey Questions

Item Number	Question	Responses
1	I worried about having a place to stay yesterday.	1.Yes 2.No
2	I could not get to where I needed to go yesterday.	1.Yes 2.No
3	I worried about my food running out yesterday.	1.Yes 2.No
4	I ate less than I needed to yesterday.	1.Yes 2.No
5	I was hungry, but did not eat because I couldn't afford food.	1.Yes 2.No
6	I got food from a pantry, church, agency, friend or the street yesterday.	1.Yes 2.No
7	Did you drink alcohol yesterday?	1.Yes 2.No
8	How many alcohol drinks did you have yesterday?	0-20

Beyond the daily questionnaire, data was also collected through a computer-assisted interview on behavioral, psychosocial, and cognitive factors (ex. Income, alcohol-toxicity beliefs, education, etc.) which have been shown to be associated with medication adherence. These values were only collected at baseline and are considered time independent. The variables from this assessment used in our analysis concerned education level, income, age, beliefs about the interactivity of alcohol with ART medication, depression, and alcohol abuse.

3.2 Simulations

The first step in the process of creating the simulated datasets was to obtain parameter estimates from Pellowski et al. [2016] and to quantify the relationships between variables. These estimated relationships were treated as the "true" relationships and **one thousand (number or no???)** samples were drawn using these parameters. The purpose of simulating these datasets was to see how well they captured the true relationship between daily medication adherence on a certain day, whether or not a participant drank that day, and their beliefs about alcohol-medication interactivity (included as a Z-transformed test score). Specifically,

Let $Y_{ij} \sim binomial(n_{ij}, p_{ij})$ be whether the jth person missed their medication dose on the ith day. Then the model of interest is the random intercept model with time effect described as follows:

$$logit(p_{ij}) = \beta_{0j} + \beta_1 DrinkYN_i + \beta_2 ZAlcTox_i + \beta_3 Day$$

where $\beta_{0j} = \gamma_0 + \alpha_j$ and $\alpha_j \sim \mathcal{N}(0, \sigma_\alpha^2)$.

Additionally, note:

$p_{ij} = 1$ when participant p did not adhere to the ART on day i, 0 otherwise.
$DrinkYN = 1$ if participant p drank on day i, 0 otherwise.
$ZAlcTox$ = participant p's Z-transformed score on an alcohol-antiretroviral interactive toxicity scale developed by Kalichman Kalichman et al. [2009].
Day = Day i of the study, 0-44.
j = participant index, 1-60.

Based on the observed data, the true values of the parameters, γ_0, β_1, β_2, β_3, and σ_α^2 were found to be -2.29, 0.28, 0.64, 0.01, and **1.0???** respectively.

Once the samples were drawn, they were transformed under the three different missingness structures. These were the Split-Form and Wave Missingness designs previously mentioned as well as a variation on the Split-Form design where a few measurement occasions (here, two days per participant) are fully observed. To clarify this difference, see the hypothetical application of these designs to a five question multi-day survey in **Table 4 (do we want the illustration here instead? or in both places?)**. As in the previous tables, "1" indicates an observed occasion while "0" indicates a missing one. We see that the Split Form design entails missing two items on each of the days. The altered Split Form, while similar, is fully measured on two of the measurement occasions. Because there are days which are fully observed, the relationships between variables may be clearer, allowing the imputation procedure to run more effectively. On the other hand, the Wave Missingness design shows

two days in which no questions are asked but the full questionnaire is given on the remaining days.

Table 4: Different Planned Missingness Designs

Design	Day	Q1	Q2	Q3	Q4	Q5
Split Form	1	1	1	0	0	1
	2	0	0	1	1	1
	3	0	1	1	0	1
	4	1	0	1	1	0
	5	1	1	0	1	1
Altered Split Form	1	1	1	1	1	1
	2	1	1	1	1	1
	3	0	0	1	1	1
	4	1	1	0	1	0
	5	0	1	1	0	1
Wave Missingness	1	1	1	1	1	1
	2	0	0	0	0	0
	3	1	1	1	1	1
	4	1	1	1	1	1
	5	0	0	0	0	0

Additionally, for each different Missingness pattern we imposed low, medium, and high levels of missingness roughly corresponding to 2, 4, and 6 missing questions. Previous work by Rhemtulla et al. [2014] found that for low levels of missingness Split Form design outperforms the Wave Missingness design in capturing latent growth curve model parameters. However, because only low-levels of missingness are tested it is unknown whether this relationship holds at moderate or high levels. By including these different levels of missingness the performance at all levels of the missingness spectrum will be measured.

Once missingness was created, Multiple Imputation techniques were applied to estimate the relationship between medication adherence and the other variables in our model. Because the missing data mechanism is controled, the data is known to be MCAR. This fact allows for the principled use of MI techniques. Using the R package {mice} [van Buuren and Groothuis-Oudshoorn, 2011] the binary variables (Q1-Q7) were imputed using a logistic model based approach while the count variable (Q8) was imputed using predictive mean matching **was it??? i don't think so, it was in the X block originally but double check!**.

After imputing, a logistic random-intercept model was estimated for each of the imputed datasets and combined using Rubin's Rules [Rubin, 2004]. From here, the results from each PM design were evaluated based on bias, percent bias, mean squared error (MSE), coverage, confidence interval (CI) length, and the fraction of missing information (FMI). The FMI is a measure of how much of the MI sample variance comes from differences between imputations. In addition to comparing the results between each of the PM designs, the same analyses were also performed before missingness was imposed.

When determining if the PM designs are performing sufficiently, it is important to consider how well the simulated samples performed before adding a missingness structure. For example, consider two full data samples with respective biases on the β_1 estimate of 0.13 and 0.15. Further, after imposing different PM design structures on these samples, the average biases are 0.14 and 0.15 respectively. It would be unfair to claim that the first PM

design performed better than the second because, by chance, it was based off of a more representative sample. This fact was considered when determining the evaluation criteria.

For bias, percent bias, and MSE we claimed that an increase of more than 40% in these measurements after imposing missingness indicates that the PM design is inappropriate. For instance, if the complete data shows a percent bias of 10%, anything higher than 14% would be unacceptable. An increase of this size has the ability to negatively effect efficiency of our estimation. Based on the work of Collins et al. [2001] we also regard coverages of less than 90% as unacceptable. If a 95% CI has a true coverage of 90% than the Type I error rate is double what it should be. Because MI accounts for uncertainty in the imputation process, coverages should be similar despite the added missingness. In terms of confidence interval lengths, if one PM design provides similar rates of coverage to another while maintaining shorter intervals than this is considered the better of the two methods. For equal coverage rates, shorter intervals imply more power. The FMI, which quantifies the uncertainty in our imputation process, is simply reported for each PM design. While less uncertainty in the imputations can be beneficial, less variation does not necessarily imply accuracy. Thus, FMI should be taken in the context of the other measurements.

This process was repeated under the three modified simulation scenarios mentioned in the following sections. Note that while typically these PM designs would be compared against the results obtained using a complete case analysis [Rhemtulla et al., 2016], this approach was avoided as it would fundamentally alter the form of the data. Specifically, complete case analysis removes any observation with any missing values. Then, for the Split Form design this would remove all observations. In the Wave Missingness setup, complete case analysis would shorten the length of the longitudinal study. If a participant was missing values on 11 of the 45 days, then complete case analysis would have made the data resemble a longitudinal study held over 34 days. Hence, this comparison is invalid.

3.2.1 Increased Sample Size

In order to evaluate the impact of sample size on estimation, the original sample size of 60 participants was expanded to 120. The rest of the simulation process was carried out as before. As noted earlier, the cost savings gleaned from the PM designs could allow us to enroll more participants. Thus, even if one design performed more optimally than the others, if all designs also perform acceptably well at high samples, researchers can design missingness in whichever method is most convenient. Furthermore, because Wave Missingness allows for greater savings, an increase in sample size is more reasonable for that type of missingness.

3.2.2 Low Inter-Survey Correlation

For our second simulation scenario we looked to determine the effect of decreased inter-survey correlation on which PM design would perform best. To clarify, consider the fact that answers given on a particular day may be correlated with one another- the two questions on housing insecurity, for example, are likely to be "yes" on the same days. While this fact was accounted for in the original simulations, it was not in this set of simulations. If the probability of answering "yes" on Question 2, for instance, was a function of the answer to Question 1 and the baseline variables in the original simulations, then in this scenario it is a function of only the baseline variables. This implies that each of the questions are conditionally independent of each other given the baseline variables. If researchers have a questionnaire where they do not expect any of the answers to be obviously related the results

here may help to choose which missingness design would be most optimal. To investigate this, the correlation of the variables was scaled by 0.5.

3.2.3 High Time Effect

In the final scenario, the goal was to determine which PM design would lead to the best estimates when time has a disproportionately large effect on our outcome variable. To do this we changed the value of the parameters γ_0, β_1, β_2, and β_3 to -4.56, 0.14, 0.32, and 0.01 respectively. Note that instead of increasing the value of β_3, the slope for time effect, all the other parameters were decreased. This was done because β_3 was already fairly large and increasing its value further would cause the probability of missing a dose to rapidly approach 1. Decreasing the other variables effects ensured that time has a disproportionate effect without this problem. The results of this section may prove useful to researchers who believe that the overall passage of time plays a larger role in their outcome than a measurement taken at one specific point.

4 Results

4.1 Primary Simulation

The first set of simulations run contained the parameters in the random-intercept logistic model with $\gamma_0 = -2.28$, $\beta_1 = 0.28$, $\beta_2 = 0.64$, and $\beta_3 = 0.01$. The effects of the variables here are comparable to those found in the original dataset. This fact implies that we can interpret the results here to show what may have occurred if the original study had implemented a PM design. Before we imposed a missingness structure on the 1,000 simulated samples we obtained the full data results shown in Table **?? not here yet**. Note that the full data results for the other simulation scenarios are also provided in this table for ease of reference.

Of note in the complete-data primary analysis is that the coverage of the 95% CIs is larger than expected for the parameter estimates and the percent bias for the β_1 estimate is somewhat large. This increased bias is likely due to the difficulties inherent in estimating logistic regression models with rare event data [King and Zeng, 2001]. Nevertheless, this must be kept in mind when analyzing the results from the PM designs provided in Table 5. **In Table 5 as in Tables 6, 7, and 8 bold values indicate "unacceptable" results.** That is, bias, percent bias, or MSE more than 40% greater than the complete-data value or a CI coverage of less than 90%.

For the Split Form and Altered Split Form designs seen in Table **??**, note that the greatest loss in efficiency was on the β_1 estimate. While the CI coverage remained high at all levels of missingness for the two designs, the bias, percent bias, and MSE reached unacceptable levels at the high missingness level. We also point out that the Altered Split Form design did not perform significantly better than the regular Split Form design. While some individual measurements outperformed those from the Split Form design, bias on β_1 for example, it underperformed in others, such as percent bias on β_3. In reference to the FMI on the β_1 estimate, we can see that at low and medium missingness both Split Form designs performed well. That is, there was less variance in the results from th imputed datasets than we expected given the amount of missing data.

Towards the bottom of Table 5, the Wave Missingness design provided unacceptable results on bias, percent bias, and MSE for the β_1 estimate at both medium and high levels

Table 5: Primary Results (N = 60)

Design	Level	Parameter	Bias	Pct Bias	MSE	Coverage	CI Length	FMI
Split Form	Low	γ_0	0.042	-1.817	0.018	0.988	0.702	0.002
		β_1	-0.071	-25.122	0.269	0.983	2.534	0.236
		β_2	0.008	1.253	0.024	0.932	0.585	0.000
		β_3	0.000	-0.886	0.000	0.944	0.017	0.001
	Medium	γ_0	0.039	-1.684	0.018	0.990	0.709	0.011
		β_1	-0.102	-36.186	0.271	0.998	3.500	0.534
		β_2	0.008	1.207	0.024	0.933	0.586	0.002
		β_3	0.000	-0.419	0.000	0.946	0.017	0.007
	High	γ_0	0.028	-1.242	0.027	0.985	0.791	0.030
		β_1	-0.107	-38.054	0.418	0.988	4.268	0.630
		β_2	0.012	1.869	0.027	0.932	0.613	0.005
		β_3	0.000	-0.471	0.000	0.946	0.020	0.020
Altered Split Form	Low	γ_0	0.042	-1.832	0.018	0.990	0.701	0.001
		β_1	-0.069	-24.401	0.283	0.983	2.508	0.218
		β_2	0.008	1.272	0.024	0.933	0.585	0.000
		β_3	0.000	-0.949	0.000	0.944	0.017	0.001
	Medium	γ_0	0.040	-1.760	0.018	0.988	0.704	0.007
		β_1	-0.117	-41.573	0.277	0.990	3.385	0.508
		β_2	0.008	1.286	0.024	0.931	0.585	0.002
		β_3	0.000	-0.986	0.000	0.945	0.017	0.005
	High	γ_0	0.036	-1.592	0.021	0.989	0.742	0.022
		β_1	-0.158	-56.111	0.394	0.992	4.336	0.662
		β_2	0.012	1.809	0.026	0.934	0.601	0.006
		β_3	0.000	-1.487	0.000	0.941	0.018	0.021
Wave Missingness	Low	γ_0	0.040	-1.765	0.018	0.988	0.703	0.005
		β_1	-0.094	-33.425	0.238	0.993	2.995	0.433
		β_2	0.008	1.251	0.024	0.934	0.585	0.001
		β_3	0.000	-0.875	0.000	0.943	0.017	0.002
	Medium	γ_0	0.028	-1.223	0.018	0.995	0.744	0.051
		β_1	-0.088	-31.121	0.240	0.992	3.497	0.718
		β_2	0.009	1.436	0.024	0.933	0.588	0.006
		β_3	0.000	-1.326	0.000	0.946	0.017	0.025
	High	γ_0	-0.004	0.194	0.020	0.995	0.962	0.224
		β_1	-0.080	-28.263	0.234	0.985	3.791	0.886
		β_2	0.011	1.700	0.024	0.938	0.607	0.032
		β_3	0.000	-3.304	0.000	0.970	0.020	0.148

of missingness. Moreover, at high levels of missingness coverage became unacceptable as coverage probabilities of the 95% CIs dropped to 88.5%. Additionally, we see that CI length for the β_1 estimate was roughly equal across all designs, despite the fact that the Wave Missingness design provided more biased estimates.

4.2 Increased Sample Size

In the next simulation scenario the sample size was increased to 120 participants. All parameter values were held the same as in the original simulation. Table **?? not here yet**, again, contains the results of the analysis before applying any PM design. The results of the full-data simulations are compared to those obtained from the PM designs, shown in Table 6.

First notice the bias of all parameters is lower than in the smaller sample size simulations. Next, the Split Form design performed similarly well as in the first. At low levels of missingness it gave acceptable results for bias, percent bias, and MSE for the β_1 and for medium levels only the MSE was unacceptable. However, the CI length is relatively far from

Table 6: Increased Sample Size Results (N = 120)

Design	Level	Parameter	Bias	Pct Bias	MSE	Coverage	CI Length	FMI
Split Form	Low	γ_0	0.013	-0.577	0.008	0.985	0.424	0.001
		β_1	-0.041	-14.633	0.134	0.975	1.690	0.183
		β_2	0.006	0.867	0.007	0.944	0.319	0.000
		β_3	0.000	-0.868	0.000	0.954	0.012	0.000
	Medium	γ_0	0.012	-0.544	0.008	0.985	0.429	0.009
		β_1	-0.073	-25.973	0.148	0.991	2.366	0.449
		β_2	0.006	0.875	0.007	0.944	0.320	0.002
		β_3	0.000	-0.832	0.000	0.953	0.012	0.002
	High	γ_0	0.004	-0.179	0.011	0.973	0.465	0.025
		β_1	-0.145	-51.616	0.304	0.983	3.281	0.649
		β_2	0.007	1.136	0.008	0.939	0.332	0.005
		β_3	0.000	0.243	0.000	0.944	0.013	0.010
Altered Split Form	Low	γ_0	0.013	-0.576	0.008	0.984	0.425	0.002
		β_1	-0.032	-11.186	0.129	0.980	1.677	0.173
		β_2	0.006	0.875	0.007	0.944	0.319	0.000
		β_3	0.000	-0.883	0.000	0.952	0.012	0.000
	Medium	γ_0	0.013	-0.553	0.008	0.983	0.426	0.005
		β_1	-0.058	-20.672	0.145	0.996	2.250	0.414
		β_2	0.006	0.879	0.007	0.944	0.319	0.001
		β_3	0.000	-0.917	0.000	0.951	0.012	0.002
	High	γ_0	0.011	-0.465	0.008	0.986	0.434	0.017
		β_1	-0.112	-39.796	0.228	0.994	3.245	0.691
		β_2	0.006	0.886	0.007	0.940	0.321	0.004
		β_3	0.000	-1.054	0.000	0.952	0.012	0.009
Wave Missingness	Low	γ_0	0.013	-0.575	0.008	0.985	0.426	0.004
		β_1	-0.064	-22.793	0.137	0.987	2.064	0.342
		β_2	0.005	0.832	0.007	0.944	0.319	0.001
		β_3	0.000	-0.906	0.000	0.952	0.012	0.001
	Medium	γ_0	0.010	-0.453	0.008	0.984	0.435	0.023
		β_1	-0.082	-29.029	0.184	0.994	2.979	0.630
		β_2	0.006	0.890	0.007	0.943	0.320	0.004
		β_3	0.000	-0.959	0.000	0.954	0.012	0.008
	High	γ_0	-0.010	0.428	0.010	0.988	0.554	0.179
		β_1	-0.059	-21.087	0.297	0.970	3.760	0.859
		β_2	0.008	1.204	0.007	0.952	0.334	0.040
		β_3	0.000	-2.426	0.000	0.955	0.014	0.105

the complete-data measurement in Table **?? not here yet**. The CI length from the complete-data simulations was 0.460 while at high levels of missingness the CI length reaches 1.088. Furthermore, the percent bias on the β_1 estimate is not seriously different from that obtained in the small sample scenario. For example, at medium levels of missingness the percent bias on β_1 is 23.86% while for the smaller sample it was 23.75%. Similarly to the Split Form design, the Altered Split Form Design also provided unacceptable results for the MSE of the β_1 estimate at medium missingness and unacceptable bias, percent bias, and MSE at high levels.

For the Wave Missingness design we continue to see the issues observed in the small sample simulations. That is, for the β_1 estimate the measurements at medium and high levels of missingness are unacceptable and a steep drop off in coverage probability occurs for the 95% CIs. In fact, at medium and high levels of missingness the coverages drop to 85.5% and 82.0%. The γ_0 parameter also continues to be estimated poorly at high levels of missing data under the Wave Missingness design.

4.3 Low Inter-Survey Correlation

For the next set of simulations sample size returned to 60 participants but the level of inter-survey correlation decreased. The results obtained from the full data simulations are provided in Table **?? not here yet** while the results from the PM designs are given in Table 7.

Table 7: Low Inter-Survey Correlation Results (N = 60)

Design	Level	Parameter	Bias	Pct Bias	MSE	Coverage	CI Length	FMI
Split Form	Low	γ_0	0.042	-1.856	0.018	0.990	0.702	0.002
		β_1	-0.067	-23.958	0.265	0.981	2.537	0.230
		β_2	0.007	1.122	0.024	0.938	0.585	0.000
		β_3	0.000	-1.108	0.000	0.940	0.017	0.001
	Medium	γ_0	0.041	-1.774	0.018	0.992	0.708	0.010
		β_1	-0.103	-36.645	0.281	0.991	3.492	0.535
		β_2	0.008	1.183	0.024	0.934	0.586	0.002
		β_3	0.000	-1.031	0.000	0.939	0.017	0.006
	High	γ_0	0.034	-1.488	0.026	0.990	0.794	0.030
		β_1	-0.148	-52.475	0.425	0.991	4.285	0.621
		β_2	0.011	1.791	0.028	0.935	0.613	0.005
		β_3	0.000	-0.870	0.000	0.941	0.020	0.020
Altered Split Form	Low	γ_0	0.043	-1.874	0.018	0.989	0.701	0.001
		β_1	-0.066	-23.566	0.258	0.984	2.535	0.224
		β_2	0.007	1.116	0.024	0.935	0.585	0.000
		β_3	0.000	-1.202	0.000	0.939	0.017	0.001
	Medium	γ_0	0.041	-1.782	0.018	0.988	0.706	0.009
		β_1	-0.096	-33.989	0.262	0.991	3.394	0.511
		β_2	0.008	1.226	0.024	0.934	0.585	0.002
		β_3	0.000	-1.150	0.000	0.940	0.017	0.005
	High	γ_0	0.036	-1.573	0.020	0.986	0.738	0.023
		β_1	-0.174	-61.685	0.380	0.991	4.350	0.669
		β_2	0.010	1.639	0.025	0.937	0.598	0.005
		β_3	0.000	-0.961	0.000	0.949	0.018	0.024
Wave Missingness	Low	γ_0	0.041	-1.790	0.018	0.989	0.703	0.005
		β_1	-0.084	-29.806	0.242	0.994	2.967	0.423
		β_2	0.007	1.141	0.024	0.936	0.585	0.001
		β_3	0.000	-1.100	0.000	0.940	0.017	0.002
	Medium	γ_0	0.031	-1.347	0.018	0.992	0.739	0.046
		β_1	-0.112	-39.818	0.246	0.988	3.491	0.722
		β_2	0.008	1.269	0.024	0.936	0.588	0.005
		β_3	0.000	-1.499	0.000	0.944	0.017	0.024
	High	γ_0	-0.003	0.133	0.022	0.996	0.985	0.237
		β_1	-0.094	-33.488	0.233	0.988	3.875	0.887
		β_2	0.010	1.616	0.025	0.941	0.608	0.033
		β_3	-0.001	-3.529	0.000	0.964	0.021	0.160

For the Split Form design it is important to notice that the efficiency of the β_1 estimate decreases much more rapidly than in the original simulations where survey questions were predictive of one another. For instance, the bias of β_1 at low levels of missingness, here, was found to be 0.150 and 0.276 at high levels. This left us with an increase in average bias of 0.126. Looking at the same increase in the original simulations, note that the bias only increased by 0.071. This loss in efficiency is seen in the other measurements on the β_1 estimate as well. Moreover, note that the Split Form design performs unacceptably in terms of bias, percent bias, and MSE at medium levels of missingness. This was not the case in the other missingness scenarios.

The Altered Split Form design at medium levels of missingness yielded an MSE outside of the acceptable range and at high levels bias, the percent bias are unacceptable as well. Additionally, the coverage of the β_1 CIs was quite a bit higher than in the full-data simulations. That is, while the coverage probability obtained in the full-data simulations was 94.5%, under this Altered Split Form design the coverages are 95.5%, 97.0%, and 96.0% at low, medium, and high levels of missingness respectively.

Under Wave Missingness the issues of unacceptable bias, percent bias and MSE occur. However, at low levels of missingness, the coverage of the β_1 estimate was fairly good (96.5% compared to 94.5% in the full data). Sufficient coverage did not hold through increased missingness as coverage dropped to 89.5% in the high missingness scenario. As in other scenarios, the CI lengths produced under the Wave Missingness design are similar to those produced in the Split Form designs despite an increase in bias. However, here the increase in bias is less dramatic than in other scenarios. For example, at high levels of missingness in the original simulations the average bias on the β_1 estimate was 0.065 higher for the Wave Missingness design (at 0.290) than in the Split Form design (0.225). In this simulation scenario that same difference was only 0.027. Moreover, while the estimation of γ_0 was unacceptable at high levels of missingness in the other simulation scenarios, here it was in the acceptable range.

4.4 High Time Effect

In the final scenario we tested to see how the PM designs would capture a model where time had a disproportionately large effect on Missed Dose. In this scenario parameter values were changed to $\gamma_0 = -4.6$, $\beta_1 = 0.28$, $\beta_2 = 0.64$, and $\beta_3 = 0.2$. The results from our full-data analyses are provided in Table ?? while the corresponding PM design results are in Table 8.

The effects of imposing Split Form missingness in this high time effect scenario are similar to those under the low inter-survey correlation scenario. At low levels of missingness bias, percent bias, MSE, and coverage for β_1 are all within the acceptable range. At medium and high levels of missingness bias, percent bias, and MSE are unacceptably high.

The Altered Split Form design performed slightly better than the normal Split Form design under this simulation scenario. Note, however, that while the normal Split Form design provided unacceptable bias and percent bias for the β_1 estimate at medium levels of missingness, this was not true of the Altered Split Form design. Additionally, observe that at medium and high levels of missingness there was an improvement under the Split Form design in the CI lengths for the β_1 estimate. In the Split Form design the CI lengths were recorded as 0.981 and 1.433 at medium and high missingness. Under the Altered Split Form design these measurements were 0.903 and 1.381.

In this simulation scenario, the Wave Missingness Design performs similarly to the Split Form and Altered Split Form designs at low levels of missingness. The bias, percent bias, MSE, and CI coverage for β_1 all performed acceptably using the Wave Missingness design at low levels of missingness. However, as missingness increases this design shows itself to be unsuitable in terms of estimating β_1 for the same reasons noted in the other scenarios.

4.5 Interpretation

We note two commonalities throughout our simulation scenarios. The first is that the Altered Split Form design consistently failed to produce better results than the normal Split Form

Table 8: High Time Results (N = 60)

Design	Level	Parameter	Bias	Pct Bias	MSE	Coverage	CI Length	FMI
Split Form	Low	γ_0	-0.023	0.498	0.120	0.968	1.414	0.018
		β_1	0.604	429.053	0.669	0.973	4.812	0.271
		β_2	0.010	3.263	0.048	0.955	0.827	0.003
		β_3	0.000	-2.671	0.000	0.945	0.039	0.004
	Medium	γ_0	-0.048	1.060	0.128	0.972	1.506	0.059
		β_1	0.724	513.940	0.898	0.988	5.933	0.499
		β_2	0.025	7.927	0.047	0.958	0.834	0.013
		β_3	0.000	-2.867	0.000	0.938	0.039	0.017
	High	γ_0	-0.178	3.882	0.370	0.984	2.050	0.132
		β_1	0.997	708.287	1.573	0.955	6.822	0.610
		β_2	0.044	13.650	0.060	0.965	0.966	0.036
		β_3	0.000	-0.787	0.000	0.955	0.048	0.060
Altered Split Form	Low	γ_0	-0.019	0.415	0.119	0.966	1.401	0.011
		β_1	0.603	428.601	0.638	0.966	4.616	0.239
		β_2	0.014	4.254	0.048	0.956	0.827	0.002
		β_3	0.000	-3.004	0.000	0.944	0.039	0.003
	Medium	γ_0	-0.045	0.988	0.124	0.978	1.493	0.053
		β_1	0.728	517.389	0.879	0.974	5.976	0.504
		β_2	0.024	7.437	0.051	0.954	0.838	0.011
		β_3	-0.001	-4.550	0.000	0.946	0.039	0.018
	High	γ_0	-0.117	2.557	0.196	0.984	1.759	0.112
		β_1	0.876	622.327	1.305	0.968	6.913	0.650
		β_2	0.026	8.265	0.051	0.963	0.892	0.034
		β_3	-0.001	-5.340	0.000	0.949	0.043	0.055
Wave Missingness	Low	γ_0	-0.057	1.249	0.127	0.968	1.454	0.038
		β_1	0.853	606.148	1.003	0.974	5.629	0.510
		β_2	0.006	1.804	0.046	0.959	0.831	0.006
		β_3	0.000	-3.051	0.000	0.945	0.039	0.011
	Medium	γ_0	-0.182	3.981	0.166	0.979	1.813	0.183
		β_1	0.979	695.352	1.325	0.967	6.758	0.782
		β_2	0.020	6.425	0.050	0.956	0.871	0.037
		β_3	-0.001	-4.466	0.000	0.955	0.042	0.068
	High	γ_0	-0.403	8.803	0.345	0.987	2.789	0.440
		β_1	1.050	745.735	1.504	0.951	7.357	0.887
		β_2	0.019	5.848	0.050	0.973	0.960	0.099
		β_3	-0.003	-18.904	0.000	0.975	0.052	0.200

design. Secondly, the Wave Missingness Design resulted in unacceptably poor estimation for anything but low levels of missingness.

This first result, the lack of improvement seen in the Altered Split form design is surprising. By keeping a small number of measurement occasions fully observed we would have expected the imputation process to perform more effectively and improve our estimates. However, even factoring in the slight decrease in overall missingness, this design performed only equally as well as the normal Split Form design. This may be a result of the fact that the number of observations we left fully observed was small; only two out of the forty-five days (approximately 4.4% of days). While it may be interesting to see if the inclusion of more fully observed days is capable of improving our estimates, perhaps keeping five days (11.1%) fully observed, we would need to be careful about raising this number too high. If the cost or burden of observations is a limiting factor, then collecting full data on even 10% of the days may be unnecessarily burdensome.

Of larger concern, however, is the steep drop-off in estimating quality of the Wave Missingness design as missingness increased. In every scenario tested, the Wave Missingness design produced unacceptable bias, percent bias, and MSE at medium and high levels of missingness. In addition, despite increased bias on estimates with missing values, the length

of confidence intervals for these estimates remained the same as in the Split Form design. This implies that there is an underestimation of the variances occurring at some point in the imputation process. This is supported by the fact that, generally, the fraction of missing information estimated in the Wave Missingness design tended to be smaller than that of the corresponding Split Form design. This can also explain the lower confidence interval coverages obtained using this design. At high levels of missingness none of the scenarios allowed the Wave Missingness design to obtain coverages above 90%. This leads to the belief that many of the imputations are producing datasets which are similar in terms of the relationships between variables but that are not representative of the true relationships. Because of this, the use of Wave Missingness designs for anything but small levels of missing data cannot be recommended.

Additionally, all PM designs under all scenarios performed within acceptable ranges when measuring the effects of variables which had no missing values ($\gamma_0, \beta_2, \beta_3$). This fact makes the increased sample size scenario promising for researchers. While our estimation of β_1 was similar in this scenario as in others, the estimation of the other model parameters improved by leaps and bounds. If a PM design is able to reduce costs by 50% by introducing missingness, using these extra funds to recruit more participants may be a good idea for researchers whose analysis of interest involves variables that will not have missing values.

5 Discussion

In this paper we performed a simulation study to test the ability of different PM designs to capture a multilevel logistic model under a longitudinal framework. We found that for low levels of planned missing data (about 25% missingness), variations of the Split Form design and the Wave Missingness designs are all capable of estimating the parameters of the multilevel model reasonably well. As the level of missing data increases to 50% and 75%, however, the quality of estimates in all of the tested PM designs become unacceptable as we see rising parameter biases and confidence interval lengths. Moreover, we found that the quality of the Wave Missingness design depreciates much more rapidly than the other PM designs, providing higher parameter biases than the corresponding Split Form estimates and low confidence interval coverages (95% confidence intervals with less than 90% coverage in some cases).

These facts warn against the use of Wave Missingness designs when the Split Form design can be easily substituted. If Wave Missingness must be used, as in the cohort sequential and developmental timelLag designs, the design must minimize the amount of missing data. For instance, in a cohort sequential design which attempts to stretch four-year data to cover a span of six years each cohort will have missing values at two of the six measurement occasions. This would result in about 33.3% missingness. To ensure that the researcher's analyses are accurate, we would recommend an increase in the amount of time they follow participants, using five-year data instead of four-year data to cover the same period. While the recommendation to use low levels of missing data is especially true for Wave Missingness designs it also holds for the Split Form designs. While these designs tended to perform well for all parameters in the multilevel model at low levels of missingness this was not true when missingness was increased to around 50%.

For future research it may be useful to test a wider range of missingness. That is, while this study showed that the PM designs performed well when missingness was around 25% and sometimes performed poorly when it was around 50% it is unknown exactly when the estimation procedure breaks down. While only 25% missing data is suggested here, if it

was shown that these designs performed well at 40% missingness, then researchers could substantially increase cost reduction and decrease participant burden beyond what our recommendations allow. Another worthwhile pursuit may be to see how hybrid PM designs perform. While it was found that Wave Missingness designs performed underwhelmingly, it would be interesting to investigatte a design containing fully missing days with the remaining days missing on some values would perform. This investigation may be particularly useful to researchers who wish to implement an accelerated longitudinal design but also have measures that are costly or burdensome which they would like to limit the collection of.

To conclude we discuss the limitations of this analysis. DIFFERENT LIMITATIONS. NOT COMPUTING POWER.

Hopefully, the results of this paper will serve to help researchers decide which PM design is most appropriate for their studies' needs. By illuminating the weaknesses of the Wave Missingness design, it is hoped that, unless it is explicitly required, researchers will opt instead to use a Split Form design. Furthermore, when researchers do implement a PM design we hope they follow our recommendation to err on the side of caution and keep the amount of missing data relatively low.

References

E Andrés Houseman and Donald K Milton. Partial questionnaire designs, questionnaire non-response, and attributable fraction: applications to adult onset asthma. *Statistics in medicine*, 25(9):1499–1519, 2006.

Linda M Collins, Joseph L Schafer, and Chi-Ming Kam. A comparison of inclusive and restrictive strategies in modern missing data procedures. *Psychological methods*, 6(4): 330, 2001.

Arthur P Dempster, Nan M Laird, and Donald B Rubin. Maximum likelihood from incomplete data via the em algorithm. *Journal of the royal statistical society. Series B (methodological)*, pages 1–38, 1977.

Terry E Duncan, Susan C Duncan, and Lisa A Strycker. *An introduction to latent variable growth curve modeling: Concepts, issues, and application*. Routledge Academic, 2013.

Amanda C Gottschall, Stephen G West, and Craig K Enders. A comparison of item-level and scale-level multiple imputation for questionnaire batteries. *Multivariate Behavioral Research*, 47(1):1–25, 2012.

John W Graham, Scott M Hofer, and David P MacKinnon. Maximizing the usefulness of data obtained with planned missing value patterns: An application of maximum likelihood procedures. *Multivariate Behavioral Research*, 31(2):197–218, 1996.

John W Graham, Bonnie J Taylor, and Patricio E Cumsille. Planned missing-data designs in analysis of change. 2001.

John W Graham, Bonnie J Taylor, Allison E Olchowski, and Patricio E Cumsille. Planned missing data designs in psychological research. *Psychological methods*, 11(4):323, 2006.

John W Graham, Allison E Olchowski, and Tamika D Gilreath. How many imputations are really needed? some practical clarifications of multiple imputation theory. *Prevention Science*, 8(3):206–213, 2007.

Ofer Harel. Inferences on missing information under multiple imputation and two-stage multiple imputation. *Statistical Methodology*, 4(1):75–89, 2007.

Ofer Harel and Xiao-Hua Zhou. Multiple imputation: review of theory, implementation and software. *Statistics in medicine*, 26(16):3057–3077, 2007.

Ofer Harel, Jeffrey Stratton, and Robert Aseltine. Designed missingness to better estimate efficacy of behavioral studies-application to suicide prevention trials. *Journal of Medical Statistics and Informatics*, 3(1):2, 2015.

Seth C Kalichman, Christina M Amaral, Denise White, Connie Swetsze, Howard Pope, Moira O Kalichman, Chauncey Cherry, and Lisa Eaton. Prevalence and clinical implications of interactive toxicity beliefs regarding mixing alcohol and antiretroviral therapies among people living with hiv/aids. *AIDS patient care and STDs*, 23(6):449–454, 2009.

Seth C Kalichman, Christina M Amaral, Denise White, Connie Swetsze, Moira O Kalichman, Chauncey Cherry, and Lisa Eaton. Alcohol and adherence to antiretroviral medications: interactive toxicity beliefs among people living with hiv. *Journal of the Association of Nurses in AIDS Care*, 23(6):511–520, 2012.

Gary King and Langche Zeng. Logistic regression in rare events data. *Political analysis*, pages 137–163, 2001.

Roderick JA Little and Donald B Rubin. *Statistical analysis with missing data*. John Wiley & Sons, 2014.

Todd D Little. *Longitudinal structural equation modeling*. Guilford Press, 2013.

Todd D Little and Mijke Rhemtulla. Planned missing data designs for developmental researchers. *Child Development Perspectives*, 7(4):199–204, 2013.

John J McArdle and Richard W Woodcock. Expanding test–retest designs to include developmental time-lag components. *Psychological Methods*, 2(4):403, 1997.

Jennifer A Pellowski, Seth C Kalichman, Moira O Kalichman, and Chauncey Cherry. Alcohol-antiretroviral therapy interactive toxicity beliefs and daily medication adherence and alcohol use among people living with hiv. *AIDS care*, 28(8):963–970, 2016.

Trivellore Raghunathan. *Missing data analysis in practice*. CRC Press, 2015.

Trivellore E Raghunathan and James E Grizzle. A split questionnaire survey design. *Journal of the American Statistical Association*, 90(429):54–63, 1995.

Mijke Rhemtulla, Fan Jia, Wei Wu, and Todd D Little. Planned missing designs to optimize the efficiency of latent growth parameter estimates. *International Journal of Behavioral Development*, page 0165025413514324, 2014.

Mijke Rhemtulla, Victoria Savalei, and Todd D Little. On the asymptotic relative efficiency of planned missingness designs. *Psychometrika*, 81(1):60–89, 2016.

D.B. Rubin. *Multiple Imputation for Nonresponse in Surveys*. Wiley Classics Library. Wiley, 2004. ISBN 9780471655749. URL `https://books.google.com/books?id=bQBtw6rx_mUC`.

Donald B Rubin. Inference and missing data. *Biometrika*, pages 581–592, 1976.

Willem E Saris, Albert Satorra, and Germà Coenders. 8. a new approach to evaluating the quality of measurement instruments: The split-ballot mtmm design. *Sociological Methodology*, 34(1):311–347, 2004.

David M Shoemaker. *Principles and procedures of multiple matrix sampling*. Ballinger, 1973.

Paul J Silvia, Thomas R Kwapil, Molly A Walsh, and Inez Myin-Germeys. Planned missing-data designs in experience-sampling research: Monte carlo simulations of efficient designs for assessing within-person constructs. *Behavior research methods*, 46(1):41–54, 2014.

Kenneth Sirotnik and Roger Wellington. Incidence sampling: An integrated theory for matrix sampling. *Journal of Educational Measurement*, 14(4):343–399, 1977.

Neal Thomas, Trivellore E Raghunathan, Nathaniel Schenker, Myron J Katzoff, and Clifford L Johnson. An evaluation of matrix sampling methods using data from the national health and nutrition examination survey. *Survey Methodology*, 32(2):217, 2006.

Stef van Buuren and Karin Groothuis-Oudshoorn. mice: Multivariate imputation by chained equations in r. *Journal of Statistical Software*, 45(3):1–67, 2011. URL `http://www.jstatsoft.org/v45/i03/`.

Sholom Wacholder, Raymond J Carroll, David Pee, and Mitchell H Gail. The partial questionnaire design for case-control studies. *Statistics in Medicine*, 13(5-7):623–634, 1994.

Linda M Zeger and Neal Thomas. Efficient matrix sampling instruments for correlated latent traits: examples from the national assessment of educational progress. *Journal of the American Statistical Association*, 92(438):416–425, 1997.