

Sankhya B

Comparing Different Planned Missingness Designs in Longitudinal Studies

--Manuscript Draft--

Manuscript Number:	
Full Title:	Comparing Different Planned Missingness Designs in Longitudinal Studies
Article Type:	Original Research Articles
Keywords:	Planned Missingness; Split Form; Multiform; Wave Missingness; Multiple Imputation; Missing Data.
Corresponding Author:	Ofer Harel University of Connecticut UNITED STATES
Corresponding Author Secondary Information:	
Corresponding Author's Institution:	University of Connecticut
Corresponding Author's Secondary Institution:	
First Author:	Nicholas Illenberger
First Author Secondary Information:	
Order of Authors:	Nicholas Illenberger
	Jennifer Pellowski, PhD
	Ofer Harel
Order of Authors Secondary Information:	
Funding Information:	
Abstract:	<p>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.</p>

Sankhya B manuscript No. (will be inserted by the editor)

Comparing Different Planned Missingness Designs in Longitudinal Studies

Nicholas Illenberger · Jennifer Pellowski · Ofer Harel

Received: date / Accepted: date

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.

Keywords Planned Missingness · Split Form · Multiform · Wave Missingness · Multiple Imputation · Missing Data.

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 in 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 at a single measurement occasion [Raghunathan and Grizzle, 1995]. The Wave Missingness design, on the other hand, would require missing out on 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 a multi-level regression model is the analysis of interest, we would like to see if it is more efficient to design missingness using the Split Form design or the Wave Missingness design. 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]. Because our model of interest is a mixed effects model, though, we can not be sure that this result will hold.

To answer our question we performed a simulation study based on data obtained in longitudinal research on HIV medication adherence and alcohol-interactivity beliefs [Pellowski et al., 2016]. For each of the PM designs mentioned we simulated sample data conforming to that particular design and then used multiple imputation (MI) [Little and Rubin, 2014] techniques to analyze our data. We subsequently evaluated the PM methodologies using 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 our results more generalizable the experiment was repeated under a few different 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 so named because of the way in which they are implemented. Instead of giving participants one large survey, researchers will create multiple forms, each containing only a subset of the full surveys questions. A common setup is the three-form design [Graham et al., 1996]. In the three-form design 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 other three blocks. This leaves us with the configurations seen in Table 1, where "1" indicates that the block is included and "0" indicates that it is excluded. We note that every block is paired together in at least one of the forms. By designing our forms in this manner we are able to estimate the relationship between any given items regardless of which blocks they are in. That is, if we want to estimate the relationship between an item in block A and another in C, there is a form that allows us to do that.

Table 1: The Three Form PM Design

Form	Block			
	X	A	B	C
1	1	1	1	0
2	1	1	0	1
3	1	0	1	1

An important consideration when designing a Split Form survey is deciding which items to put into which blocks. The X-block, being included in every form, should contain either those variables which are 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]. For the same reason it is good practice to place variables which are predictive of each other in to separate blocks [Gottschall et al., 2012]. We note that when groupings are not as obvious as in the previous example pilot studies can help determine the predictive potential of variables. We may also wish to consider the cost and burden associated with an item when deciding which block to place it in. Even if a specific item is critical to our analysis, if it is overly expensive or invasive we may be better off excluding it from the X-block [Little and Rhemtulla, 2013].

More than deciding which blocks should contain which items, we also need to consider how many blocks we should use. Including more blocks in our design can allow us to further increase missingness [Graham et al., 2006]. For example, imagine two groups of researchers, the first group divides their survey into an X-block and three other blocks of equal size, each containing 25% of the survey, and the second uses an X-block and nine others with 10% of the items each. The researchers make the decision that any given pair of blocks should be asked together in at least one of the forms to ensure efficient estimations. Based on this requirement, the first group can only reduce the length of their survey by 25% while the second group can decrease the length of their survey by up to 70%.

While this increase in missingness may allow for further reductions in burden and cost researchers must consider the trade-offs of using more blocks. Besides the fact that greater missingness will lead to less efficient estimation [Rhemtulla et al., 2016] there is the added complication that more blocks means more forms. If the second group of researchers in our example, for instance, did decide that each form would contain the X-block and two of the other blocks, then they would be left with 36 total forms. For traditional paper surveys this can make increasing the number of blocks unwieldy. However, for electronic surveys, where form randomization can be automated, this is not an issue. Oftentimes, Researchers using electronic surveys will give each participant a random sample of all the survey items, in a way treating each item as its own block [Silvia et al., 2014]. Using this method researchers can reap the benefits of 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 different 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 were assigned to miss a subset of items in a single measurement occasion, the Wave Design entails randomizing participants to miss an entire measurement occasion. An example of a Wave Missing design in which data is collected every month for five months but each participant is only measured at three of these occasions as shown in Table 2. Again, "1" indicates that a given wave is included while "0" indicates that it is excluded. Subject two, for instance, is randomly assigned to be measured only at the first, second, and fourth month.

A relative advantage of the Wave Missingness design over the Split form design is that researchers can have more 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. The data described by Pellowski et al. [2016]. Pellowski et al. [2016] for instance has daily 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 have the ability to eclipse Split Form designs in terms of cost reduction [Little and Rhemtulla, 2013].

Another advantage of this design is that many common longitudinal study designs already make use of the Wave Missingness structure [Little and Rhemtulla, 2013]. Cohort Sequential Designs are a type of study where cohorts of different ages are recruited and followed longitudinally in such a way that we can use information from one cohort to impute values for another [Duncan et al., 2013, Little, 2013]. Say, for example, we had three cohorts of ages 10, 11, and 12 at baseline and we measured them 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

Table 2: The Wave Missingness PM Design

Participant	Month				
	1	2	3	4	5
1	1	0	1	0	1
2	1	1	0	1	0
3	0	0	1	1	1
4	1	1	1	0	0

Table 3: 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

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 we use it to impute missing values for the other groups. By continuing in this manner we are able to take the data that was collected over a four year period and extend it to cover six years, accelerating the study. The setup of this type of study is demonstrated in Table 3, where "1" and "0" respectively indicate an included and excluded wave.

Developmental time-lag models [McArdle and Woodcock, 1997] also make use of the Wave Missingness structure. As shown in Table 4, researchers here 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 are able to use missing data techniques to 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 we can accomplish this task with only two measurement occasions per participant. These designs exemplify how useful the Wave Missingness structure can be in accelerating longitudinal studies.

2.3 Multiple Imputation

2.3.1 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 supposed 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

Table 4: 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

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]

2.3.2 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 some 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 these datasets have been created, we can run the analysis of interest on each of them and, using rules defined by Little and Rubin [2014] and Rubin [2004], combine our 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 $\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_M$ be the estimates of θ and let S_1, S_2, \dots, 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,

$$\hat{\theta} = \sum_{n=1}^M \hat{\theta}_n. \quad (1)$$

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

Table 5: 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

$$T_M = W_M + (1 + 1/M) * B_M. \quad (2)$$

These equations form the basis of Rubin's combining rules and should provide parameter estimates which are unbiased while 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 speaking, the lower the level of missingness is the fewer imputations are needed. If less than 20% of the requisite data is missing, for instance, ten or even five 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 we based our simulations on came from a study by Pellowski et al. [2016] and more in depth information on it can be found there. However, to give a brief overview, the data was collected in an attempt to determine the day-level relationship between alcohol use, interactive toxicity beliefs, and adherence to antiretroviral therapy (ART) amongst patients with HIV. Previous studies noted 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 had 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 days and sending them a daily questionnaire. Questions in the survey concerned livelihood insecurities from the previous day as well as alcohol use. The analysis used 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 5.

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 our simulated datasets was to use the data from Pellowski et al. [2016] and obtain parameter estimates to quantify the relationships between variables. We then treated these estimated relationships as the true relationships and drew two-hundred samples using these parameters. While we would have liked to simulate a larger number of samples we were limited by computational ability. Our purpose in simulating these datasets was to see how well they would be able to capture the true relationship between 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, our model of interest is the random intercept model with time effect found below:

$$\begin{aligned} \text{logit}(\text{MissedDose}) &= \beta_{0j} + \beta_1 * \text{DrinkYN} + \beta_2 * \text{ZAlcTox} + \beta_3 * \text{Day} \\ \text{Level2} : \beta_{0j} &= \gamma_0 + \epsilon_j \end{aligned}$$

where,

MissedDose = 1 if the participant did not adhere to the ART, 0 otherwise.

DrinkYN = 1 if the participant drank on a given day, 0 otherwise.

ZAlcTox = Z-transformed score on an alcohol-antiretroviral interactive toxicity scale developed by Kalichman et al. [2009].

Day = The number of days since the study began, 0-44.

j = participant index, 1-60.

$\epsilon_j \sim N(0, \sigma^2)$.

The true values of the parameters, $\gamma_0, \beta_1, \beta_2, \beta_3$, and σ^2 were set to -2.3, 0.56, 1.28, 0.2, and 1.0 respectively.

Once our samples were drawn we imposed three different missingness structures on each of the samples corresponding to three possible PM designs. 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 6. 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

Table 6: 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

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.

Additionally, for each different Missingness pattern we imposed low, medium, and high levels of missingness roughly corresponding to 25%, 50%, and 75% missing data. Previous work by Rhemtulla et al. [2014] has found that for low levels of missingness (25%) the Three-Form design outperforms the Wave Missingness design in its ability to capture latent growth curve model parameters. However, because only low-levels of missingness are tested we do not know if this relationship holds at moderate or high levels. We hope that by including these different levels of missingness we will be able to see how these designs perform at all levels of the missingness spectrum.

Once missingness was created we used Multiple Imputation techniques to estimate the relationship between Medication Adherence and the other variables in our model. Because we have control over the missing data mechanism, we know the data is MCAR. This fact allows for the principled use of MI techniques. Using the R package `{mice}` [van Buuren and Groothuis-Oudshoorn, 2011] our binary variables (Q1-Q7) were imputed using a logistic model based approach while the count variable (Q8) was imputed using predictive mean matching.

After imputations the 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 sampling variance comes from differences between imputations. Besides comparing the results between each of the PM designs we also included the results of these same analyses taken from before missingness was imposed.

When determining if the PM designs are performing sufficiently well it is important to consider how well the simulated samples performed before we added a missingness structure. For example, imagine we have 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 we obtained the average biases of 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. We kept this fact in mind when determining our 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% ($0.1 * 1.4$) 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% then the Type I error rate is double what it should be. Because MI accounts for uncertainty in the imputation process coverages should have been 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.

We repeated this process under three modified simulation scenarios mentioned in the following sections. We note that while typically these PM designs would be compared against the results obtained using a complete case analysis [Rhemtulla et al., 2016] we decided against this approach because it would fundamentally alter the form of our data. To be more concrete, complete case analysis on the Split Form design would remove all observations (all observations have some missing values). 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 we took the format of the original data, which included sixty participants, and expanded it to include one-hundred and twenty. The rest of the simulation process was carried out as before. We noted earlier that cost savings created 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, then researchers can design missingness in whichever method is most convenient. We further note, that 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.

3.2.3 High Time Effect

In this final scenario we hoped 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 $\gamma_0, \beta_1, \beta_2$, and β_3 to -4.6, 0.28, 0.64, and 0.2 respectively. We note that instead of increasing the value of β_3 , the slope for time effect, we decreased all the other parameters. We did this because β_3 was already fairly large and increasing its value further would cause the probability of missing a dose to rapidly approach 1. By decreasing the other variables effects we were able to ensure that time has a disproportionate effect without incurring 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 we ran had the parameters in our random-intercept logistic model set as $\gamma_0 = -2.3$, $\beta_1 = 0.56$, $\beta_2 = 1.28$, and $\beta_3 = 0.2$. 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 our two-hundred simulated samples we obtained the full data results shown in Table 7. We note that the full data results for the other simulation scenarios are also provided in this table for ease of reference.

Table 7: Results of Analysis before imposing Missingness in each Scenario

Scenario	Parameter	Bias	Percent Bias	MSE	Coverage	CI Length
Primary Analysis (N=60)	γ_0	0.151	6.55%	0.037	99.0%	0.750
	β_1	0.133	23.75%	0.028	95.5%	0.652
	β_2	0.120	9.40%	0.022	97.5%	0.610
	β_3	0.007	3.42%	7.11×10^{-5}	97.0%	0.034
Increased Sample (N=120)	γ_0	0.106	4.61%	0.018	98.0%	0.528
	β_1	0.099	17.63%	0.016	93.0%	0.460
	β_2	0.091	7.12%	0.013	95.5%	0.429
	β_3	0.005	2.34%	3.52×10^{-5}	96.5%	0.024
Low Inter-Survey Cor- relation (N=60)	γ_0	0.145	6.29%	0.033	99.0%	0.747
	β_1	0.130	23.27%	0.027	94.5%	0.651
	β_2	0.126	9.83%	0.026	94.0%	0.611
	β_3	0.007	3.69%	8.88×10^{-5}	93.0%	0.034
High Time Effect (N=60)	γ_0	0.170	3.69%	0.045	98.5%	0.888
	β_1	0.123	43.86%	0.024	96.0%	0.598
	β_2	0.124	19.35%	0.025	93.5%	0.575
	β_3	0.006	2.95%	5.50×10^{-5}	95.5%	0.030

Of note in the complete-data primary analysis is that the coverage of our 95% CIs is larger than expected for our 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, we should keep this in mind when analyzing our results from the PM designs provided in Table 8. In Table 8 as in Tables 9, 10, and 11 bold values indicate "Unacceptable" results. That is, a 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 8, we 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 to say, there was less variance in the results from our imputed datasets than we expected given the amount of missing data.

Towards the bottom of Table 8 we see that the Wave Missingness design provided unacceptable results on bias, percent bias, and MSE for the β_1 estimate at medium and high levels of missingness. Moreover, at high levels of missingness even 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

The next simulation scenario we tested was under an increased sample size. All parameter values were held the same as in the original set of simulations but instead of simulating data for 60 participants here we used 120 participants. Table 7, again, contains the results of our analysis before applying any PM design.

We compare the results from the full-data simulations to those obtained from the PM designs, shown in Table 9. The first thing we notice is that the bias on all parameters is lower than it was in the smaller sample size simulations. Next, the Split Form design performed similarly well under this simulation scenario 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 here is relatively far from the complete-data measurement in Table 7. The CI length from the complete-data simulations was 0.460 while at high levels of missingness the CI length reaches 1.088. We also point out that 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 here 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 pointed out in the small sample simulations. That is, for the β_1 estimate we have unacceptable measurements at medium and high levels of missingness and a steep drop off in coverage probability for our

95% CIs. In fact, at medium and high levels of missingness we find coverages of 85.5% and 82.0%. We also continue to see that the γ_0 parameter is being 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 we returned to the smaller sample size but decreased the level of inter-survey correlation. As mentioned in section 3.2.2 we made sure that there was no direct predictive relationship between questions in our survey. The results obtained from our full data simulations are provided in Table 7 while the results from our PM designs are given in Table 10.

For the Split Form design it is important to notice that the efficiency of the β_1 estimate decreases much more rapidly than in our first set of 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. If we check the same increase in the original simulations we see 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, we 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.

In the Altered Split Form design at medium levels of missingness the MSE is outside of the acceptable range and at high levels bias, are percent bias are unacceptable as well. We further note that 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 we obtained coverages of 95.5%, 97.0%, and 96.0% at low, medium, and high levels of missingness respectively.

Concerning the Wave Missingness design we, again, see the issues of unacceptable bias, percent bias and MSE. However, at low levels of missingness the coverage of our β_1 estimate was fairly good (96.5% compared to 94.5% in the full data). This positive coverage did not hold throughout the other levels of 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, this 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. 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 7 while the corresponding PM design results are in Table 11.

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. We do note, though, 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 in the Altered Split Form design. Additionally, we point out that at medium and high levels of missingness we saw an improvement from 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.

Under this simulation scenario, at low levels of missingness we see that the Wave Missingness Design performs similarly to the Split Form and Altered Split Form designs. The bias, percent bias, MSE, and CI coverage for β_1 all performed acceptably using the Wave Missingness design at low levels of missingness. Moreover, 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 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 (or about 4.4% of days) were fully observed. 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, though, is the steep drop-off in estimating quality of the Wave Missingness design as missingness increased. In every scenario we tested, the Wave Missingness design produced unacceptable bias, percent bias, and MSE at medium and high levels of missingness. We also note that 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 our 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 low confidence interval coverages we 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 us to believe 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 we cannot recommend the use of Wave Missingness designs for anything but small levels of missing data.

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 who's 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%, though, 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 lead us to recommend 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 Time-Lag designs, we suggest that the design minimizes 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 our study showed that the PM designs performed well when missingness was around 25% and sometimes performed poorly when it was around 50% we do not know exactly when the estimation procedure breaks down. While we can only recommend 25% missing data 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 we found that Wave Missingness designs performed underwhelmingly it would be interesting to see how a design which had some days fully missing and the remaining days missing on some values would have performed. In particular, looking into this result may be 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. The results we obtained from our PM designs were obtained through multiple imputation with five imputations. While this may be sufficient for low levels of missingness, it is not appropriate for some of the high levels of missingness that we tested. We chose to use such a low number of imputations because of the size of the data we had and our computational limitations. Simulating two hundred samples, each with ten variants, and then imputing five completed datasets for every variant with a missing value resulted in 9200 datasets. Then, under the four different simulation scenarios this number grew to 36,800 datasets. Factoring in the fact that each of these datasets contained at least 2700 observations (60 participants for 45 days, more for the large sample scenario) we were left with more than one-hundred million rows of data to analyze. This also led to another of the limitations we faced, using only two hundred simulated samples. While using more simulations would have allowed us to obtain more stable estimates on bias or coverage, adding in more would have added to our computational difficulties. In the future, this problem could be alleviated through the use of parallel computing through a computer cluster.

We hope that 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 we hope that when Wave missingness is not explicitly required researchers will opt instead to use a Split Form design. Moreover, 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.

Table 8: Primary Simulation Results (N=60)

Design	% Missingness	Parameter	Bias	% Bias	MSE	Coverage	CI Length	FMI
Split Form	Low(25%)	γ_0	0.151	6.59%	0.037	99.5%	0.753	0.009
		β_1	0.139	24.78%	0.029	96.0%	0.693	0.112
		β_2	0.120	9.41%	0.022	97.0%	0.610	0.002
		β_3	0.007	3.42%	7.12×10^{-5}	98.0%	0.034	0.003
	Medium(50%)	γ_0	0.153	6.67%	0.037	99.0%	0.765	0.040
		β_1	0.162	28.93%	0.041	95.5%	0.869	0.390
		β_2	0.121	9.51%	0.022	97.0%	0.616	0.007
		β_3	0.007	3.40%	6.99×10^{-5}	98.5%	0.034	0.009
	High(75%)	γ_0	0.154	6.70%	0.039	98.5%	0.818	0.147
		β_1	0.225	40.16%	0.076	97.5%	1.464	0.710
		β_2	0.119	9.28%	0.022	97.5%	0.611	0.022
		β_3	0.007	3.42%	7.02×10^{-5}	98.0%	0.035	0.032
Altered Split Form	Low(23.9%)	γ_0	0.151	6.57%	0.037	98.0%	0.753	0.008
		β_1	0.137	24.55%	0.030	96.0%	0.690	0.105
		β_2	0.120	9.39%	0.022	97.5%	0.611	0.002
		β_3	0.007	3.42%	7.10×10^{-5}	98.0%	0.034	0.003
	Medium(47.8%)	γ_0	0.153	6.67%	0.038	99.5%	0.760	0.030
		β_1	0.151	26.93%	0.036	96.0%	0.811	0.324
		β_2	0.119	9.28%	0.022	96.0%	0.611	0.006
		β_3	0.007	3.44%	7.10×10^{-5}	98.0%	0.034	0.007
	High(71.7%)	γ_0	0.164	7.13%	0.042	98.0%	0.799	0.105
		β_1	0.225	40.14%	0.080	96.0%	1.265	0.640
		β_2	0.117	9.13%	0.021	97.5%	0.615	0.019
		β_3	0.007	3.46%	7.07×10^{-5}	98.0%	0.034	0.023
Wave Missingness	Low(24.4%)	γ_0	0.152	6.61%	0.037	99.5%	0.757	0.023
		β_1	0.147	26.25%	0.033	96.0%	0.767	0.270
		β_2	0.119	9.27%	0.021	96.0%	0.610	0.004
		β_3	0.007	3.41%	6.97×10^{-5}	98.0%	0.034	0.006
	Medium(48.9%)	γ_0	0.157	6.84%	0.039	100.0%	0.767	0.052
		β_1	0.190	33.95%	0.058	94.5%	0.934	0.465
		β_2	0.119	9.27%	0.021	97.5%	0.609	0.008
		β_3	0.007	3.34%	6.80×10^{-5}	98.0%	0.034	0.010
	High(73.3%)	γ_0	0.168	7.30%	0.045	99.5%	0.807	0.129
		β_1	0.290	51.77%	0.132	88.5%	1.335	0.680
		β_2	0.119	9.32%	0.022	96.0%	0.614	0.017
		β_3	0.007	3.40%	6.96×10^{-5}	98.5%	0.034	0.025

Table 9: Increased Sample Size simulation Results (N=120)

Design	% Missingness	Parameter	Bias	% Bias	MSE	Coverage	CI Length	FMI
Split Form	Low(25%)	γ_0	0.106	4.60%	0.017	98.5%	0.533	0.022
		β_1	0.106	18.96%	0.018	96.0%	0.548	0.263
		β_2	0.091	7.14%	0.013	96.0%	0.429	0.004
		β_3	0.005	2.33%	3.47×10^{-5}	96.5%	0.024	0.005
	Medium(50%)	γ_0	0.112	4.85%	0.019	98.0%	0.542	0.057
		β_1	0.134	23.86%	0.027	95.0%	0.696	0.489
		β_2	0.092	7.15%	0.013	96.5%	0.430	0.009
		β_3	0.005	2.31%	3.46×10^{-5}	96.0%	0.024	0.011
	High(75%)	γ_0	0.110	4.80%	0.019	99.5%	0.575	0.153
		β_1	0.194	34.59%	0.057	94.5%	1.088	0.742
		β_2	0.092	7.20%	0.013	95.0%	0.432	0.021
		β_3	0.005	2.31%	3.42×10^{-5}	97.5%	0.024	0.026
Altered Split Form	Low(23.9%)	γ_0	0.107	4.64%	0.017	98.0%	0.533	0.022
		β_1	0.111	19.80%	0.020	94.5%	0.545	0.263
		β_2	0.091	7.08%	0.013	96.0%	0.429	0.004
		β_3	0.005	2.31%	3.46×10^{-5}	96.0%	0.024	0.005
	Medium(47.8%)	γ_0	0.115	4.99%	0.020	98.0%	0.542	0.056
		β_1	0.138	24.65%	0.028	94.5%	0.693	0.494
		β_2	0.091	7.24%	0.013	96.0%	0.430	0.008
		β_3	0.005	2.34%	3.54×10^{-5}	96.0%	0.024	0.011
	High(71.7%)	γ_0	0.115	5.01%	0.020	99.0%	0.563	0.119
		β_1	0.188	33.65%	0.058	91.5%	0.958	0.681
		β_2	0.092	7.16%	0.013	96.0%	0.432	0.018
		β_3	0.005	2.30%	3.44×10^{-5}	96.0%	0.024	0.020
Wave Missingness	Low(24.4%)	γ_0	0.108	4.72%	0.018	98.0%	0.533	0.024
		β_1	0.123	21.90%	0.024	92.0%	0.551	0.287
		β_2	0.092	7.18%	0.013	96.0%	0.429	0.004
		β_3	0.005	2.35%	3.51×10^{-5}	96.0%	0.024	0.005
	Medium(48.9%)	γ_0	0.111	4.84%	0.019	99.0%	0.540	0.052
		β_1	0.168	30.02%	0.043	85.5%	0.671	0.489
		β_2	0.092	7.22%	0.013	95.0%	0.429	0.007
		β_3	0.005	2.32%	3.42×10^{-5}	97.0%	0.024	0.009
	High(73.3%)	γ_0	0.123	5.33%	0.023	99.0%	0.558	0.108
		β_1	0.248	44.26%	0.099	82.0%	0.894	0.666
		β_2	0.094	7.36%	0.014	96.0%	0.430	0.013
		β_3	0.005	2.36%	3.53×10^{-5}	96.0%	0.024	0.015

Table 10: Low Inter-Survey Correlation Simulation Results (N=60)

Design	% Missingness	Parameter	Bias	% Bias	MSE	Coverage	CI Length	FMI
Split Form	Low(25%)	γ_0	0.146	6.34%	0.033	99.0%	0.755	0.025
		β_1	0.150	26.82%	0.034	95.5%	0.782	0.278
		β_2	0.126	9.82%	0.026	94.0%	0.612	0.004
		β_3	0.007	3.64%	8.68×10^{-5}	92.5%	0.034	0.007
	Medium(50%)	γ_0	0.148	6.43%	0.034	99.0%	0.772	0.067
		β_1	0.183	32.65%	0.053	95.0%	1.032	0.526
		β_2	0.125	9.80%	0.026	94.0%	0.613	0.010
		β_3	0.007	3.63%	8.73×10^{-5}	92.5%	0.035	0.015
	High(75%)	γ_0	0.155	6.74%	0.036	99.5%	0.825	0.166
		β_1	0.276	49.31%	0.116	95.0%	1.567	0.749
		β_2	0.124	9.71%	0.025	94.0%	0.618	0.025
		β_3	0.007	3.61%	8.73×10^{-5}	93.5%	0.035	0.034
Altered Split Form	Low(23.9%)	γ_0	0.141	6.13%	0.031	99.0%	0.754	0.023
		β_1	0.139	25.80%	0.030	95.5%	0.772	0.260
		β_2	0.126	9.82%	0.026	94.0%	0.612	0.004
		β_3	0.007	3.61%	8.55×10^{-5}	92.5%	0.034	0.006
	Medium(47.8%)	γ_0	0.149	6.48%	0.035	99.5%	0.767	0.056
		β_1	0.170	30.44%	0.044	97.0%	0.968	0.486
		β_2	0.125	9.79%	0.025	93.5%	0.612	0.009
		β_3	0.007	3.64%	8.58×10^{-5}	92.5%	0.034	0.011
	High(71.7%)	γ_0	0.149	6.47%	0.036	98.5%	0.809	0.139
		β_1	0.228	40.70%	0.088	96.0%	1.425	0.708
		β_2	0.125	9.74%	0.026	94.0%	0.617	0.021
		β_3	0.007	3.62%	8.65×10^{-5}	93.5%	0.035	0.027
Wave Missingness	Low(24.4%)	γ_0	0.138	6.02%	0.031	99.5%	0.754	0.025
		β_1	0.158	28.17%	0.036	96.5%	0.776	0.286
		β_2	0.126	9.82%	0.026	94.0%	0.611	0.004
		β_3	0.007	3.61%	8.50×10^{-5}	92.5%	0.034	0.006
	Medium(48.9%)	γ_0	0.148	6.43%	0.034	99.5%	0.765	0.053
		β_1	0.197	35.14%	0.059	91.0%	0.939	0.472
		β_2	0.126	9.83%	0.026	93.5%	0.611	0.007
		β_3	0.007	3.60%	8.48×10^{-5}	92.0%	0.034	0.010
	High(73.3%)	γ_0	0.158	6.86%	0.040	98.5%	0.799	0.122
		β_1	0.303	54.18%	0.148	89.5%	1.290	0.676
		β_2	0.127	9.94%	0.027	95.5%	0.615	0.015
		β_3	0.007	3.62%	8.77×10^{-5}	93.5%	0.035	0.021

Table 11: High Time Effect Simulation Results (N=60)

Design	% Missingness	Parameter	Bias	% Bias	MSE	Coverage	CI Length	FMI
Split Form	Low(25%)	γ_0	0.169	3.67%	0.045	99.0%	0.895	0.015
		β_1	0.133	47.54%	0.029	96.0%	0.708	0.262
		β_2	0.124	19.33%	0.025	93.5%	0.576	0.003
		β_3	0.006	2.96%	5.50×10^{-5}	94.5%	0.030	0.003
	Medium(50%)	γ_0	0.174	3.77%	0.047	98.5%	0.912	0.050
		β_1	0.181	64.59%	0.049	96.0%	0.981	0.554
		β_2	0.124	19.44%	0.025	94.0%	0.578	0.007
		β_3	0.006	2.97%	5.55×10^{-5}	96.0%	0.030	0.009
	High(75%)	γ_0	0.177	3.85%	0.049	98.0%	0.954	0.117
		β_1	0.243	86.63%	0.097	93.0%	1.433	0.729
		β_2	0.124	19.37%	0.025	94.0%	0.581	0.016
		β_3	0.006	2.97%	5.70×10^{-5}	94.0%	0.030	0.025
Altered Split Form	Low(23.9%)	γ_0	0.171	3.72%	0.046	99.0%	0.895	0.015
		β_1	0.142	50.55%	0.033	94.5%	0.717	0.272
		β_2	0.124	19.36%	0.025	93.5%	0.577	0.002
		β_3	0.006	2.95%	5.48×10^{-5}	95.5%	0.030	0.003
	Medium(47.8%)	γ_0	0.176	3.82%	0.048	99.0%	0.905	0.038
		β_1	0.168	60.10%	0.043	95.5%	0.903	0.486
		β_2	0.124	19.36%	0.025	93.5%	0.581	0.006
		β_3	0.006	2.95%	5.60×10^{-5}	95.0%	0.030	0.007
	High(71.7%)	γ_0	0.177	3.84%	0.051	99.5%	0.947	0.108
		β_1	0.257	91.96%	0.097	94.5%	1.381	0.714
		β_2	0.124	19.34%	0.025	94.5%	0.581	0.014
		β_3	0.006	3.01%	5.76×10^{-5}	95.0%	0.030	0.023
Wave Missingness	Low(24.4%)	γ_0	0.170	3.70%	0.046	99.0%	0.895	0.016
		β_1	0.135	48.11%	0.026	96.5%	0.720	0.290
		β_2	0.124	19.40%	0.025	93.5%	0.576	0.002
		β_3	0.006	2.94%	5.50×10^{-5}	95.0%	0.030	0.003
	Medium(48.9%)	γ_0	0.173	3.77%	0.047	99.0%	0.904	0.035
		β_1	0.186	66.53%	0.052	92.0%	0.864	0.466
		β_2	0.124	19.40%	0.025	93.5%	0.576	0.004
		β_3	0.006	2.94%	5.50×10^{-5}	96.0%	0.030	0.006
	High(73.3%)	γ_0	0.188	4.09%	0.055	98.5%	0.946	0.105
		β_1	0.322	115.10%	0.149	85.5%	1.306	0.712
		β_2	0.127	19.87%	0.026	93.5%	0.580	0.011
		β_3	0.006	3.02%	5.75×10^{-5}	95.0%	0.030	0.025