Published in final edited form as:

J Stat Plan Inference. 2016 September ; 176: 33–51. doi:10.1016/j.jspi.2016.04.001.

Multiple Imputation in Three or More Stages

J. McGinniss^{a,*} and O. Harel^b

^aBoehringer Ingelheim Pharmaceuticals, Ridgefield, Connecticut, U.S.A.

^bDepartment of Statistics, University of Connecticut, Storrs, Connecticut, U.S.A.

Abstract

Missing values present challenges in the analysis of data across many areas of research. Handling incomplete data incorrectly can lead to bias, over-confident intervals, and inaccurate inferences. One principled method of handling incomplete data is multiple imputation. This article considers incomplete data in which values are missing for three or more qualitatively different reasons and applies a modified multiple imputation framework in the analysis of that data. Included are a proof of the methodology used for three-stage multiple imputation with its limiting distribution, an extension to more than three types of missing values, an extension to the ignorability assumption with proof, and simulations demonstrating that the estimator is unbiased and efficient under the ignorability assumption.

Keywords

Multiple imputation; missing data; ignorability

1. Introduction

Incomplete data is a common obstacle to the analysis of data in a variety of fields, ranging from clinical trials to social sciences [14]. Missing values can occur for several different reasons including failure to answer a survey question, dropout, planned missing values, intermittent missed measurements, latent variables, and equipment malfunction. In fact, many studies will have more than just one type of missing value. Appropriately handling missing values is critical in the inference for a parameter of interest [3, 5]. Many methods of handling missing values inappropriately fail to account for the uncertainty due to missing values [27, 8]. This failure to account for uncertainty can lead to biased estimates and over-confident inferences.

Multiple imputation is one method for handling incomplete data that accounts for the variability of the incomplete data. This procedure does so by filling in plausible values

* née Boyko, jennifer.mcginiss@boehringer-ingelheim.com (J. McGinniss).
ofer.harel@uconn.edu (O. Harel)

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

several times to create several complete data sets and then appropriately combining complete data estimates using specific combining rules [18]. This method is praised for the ability to use complete data analytical methods on each data set as well as for retaining the variability seen in the observed data to arrive at estimates which are not distorted by the imputation method [21, 23, 10, 26].

In practice, there is often more than one type of missing value in a study. Knowledge about the nature of the missing values can help identify the most appropriate method for dealing with missing data [13]. Typically, all of the missing values are treated as though they are the same type. However, there are benefits to treating each of these types of missing values separately. One benefit arises when imputing one type of missing value first computationally simplifies the imputation of the rest of the missing values [24]. A second important benefit is that treating the two types differently can allow the researcher to quantify how much variability and how much missing information is due to each type of missing value [6]. A third benefit is the ability to assign different missing data assumptions to each of the missing value types [9]. That is, each of the missing value types might have different assumptions on the mechanisms generating the missingness. Two-stage multiple imputation is a nested version of multiple imputation where missing values of one type are imputed first [19, 11]. For each imputation of the first type of missing value, additional imputations are created for the second type of missing value, holding the first imputed values fixed. Separate combining rules determine how estimates should be handled to lead to valid inferences [24].

One area which is still unexplored is the situation where there are three or more types of missing values in a study. This is a natural extension of two-stage multiple imputation which allows for the flexibility required when dealing with real-world data. Studies which tend to have large amounts of missing values also tend to have missing values of several different types. One practical example is that of a longitudinal clinical trial where there may be missing values due to a patient being lost to follow-up, a patient missing visits intermittently over the course of a trial, and corrupted (and therefore, missing) laboratory measurements. A second example which is more relevant to surveys would be missing values due to item nonresponse, unit nonresponse, and a missing latent class. Development of a three-stage multiple imputation approach is beneficial in analyzing both of these types of studies. Three-stage multiple imputation also extends the benefits of two-stage multiple imputation, namely the quantification of the variability attributable to each type of missing value and the flexibility for greater specificity regarding data analysis. This work seeks to better simulate the reality of data collection in practice. Missing data are rarely straightforward enough to be of one, or even two, types. Multiple imputation in three stages has wide-ranging applicability across a diverse group of disciplines, as the issue of incomplete data is one which plagues researchers of all types.

The main aim of this paper is to develop the methodology for implementation of three-stage multiple imputation. In theory, the imputation stage of implementation is a simple extension of two-stage multiple imputation. However, the combining rules required for drawing appropriate inferences are different and one goal of this paper is to derive the necessary combining rules. Section 2 of this article provides an overview of standard multiple imputation and two-stage multiple imputation practices. Section 3 describes the

methodology for implementation of multiple imputation in three stages and provides a proof of the limiting distribution for the proposed estimator. Section 4 gives an extension of the methodology to k -stage multiple imputation. Section 5 defines some common missing data terminology related to mechanisms of missingness and ignorability and presents an extension to the general ignorability assumption for three types of missing values. Section 6 demonstrates use of multiple imputation in three stages with simulations, showing that the proposed estimator is unbiased and efficient, and Section 7 discusses further extensions of the methodology.

2. Multiple Imputation Review

2.1. Standard Multiple Imputation

The idea behind multiple imputation is to fill in plausible values for the missing data several times to account for model uncertainty [18, 10]. After creating $m > 1$ complete data sets by drawing from the posterior predictive distribution of the missing values, each data set is analyzed using complete data analysis methods. Let Q denote the parameter of interest. An example of such a Q might be a mean or a regression coefficient. From the complete data analyses, complete data estimates ($\hat{Q}$) and their associated variances (U) are obtained.

Let $Y = (Y_{obs}, Y_{mis})$ be the complete data where Y_{obs} is the observed part of the data and Y_{mis} is the missing part of the data. The original derivations for the combining rules were based on large sample inference [18]. Reiter and Raghunathan [16] review the implications of basing the derivation on large sample inference. The assumption involved was that, in the presence of the complete data, intervals and tests would be based on a normal approximation. That is,

$$(\hat{Q} - Q) / \sqrt{U} \sim N(0, 1).$$

The overall estimate of Q is

$$\bar{Q} = m^{-1} \sum \hat{Q}^{(j)}, \quad j=1, 2, \dots, m,$$

where $\hat{Q}^{(j)}$ is the estimate from the j^{th} repeated imputation. To get the variance for $\bar{Q}$, the between-imputation variance and the within-imputation variance must be appropriately combined. The between-imputation variance is denoted by B and is

$$B = (m-1)^{-1} \sum \left(\hat{Q}^{(j)} - \bar{Q} \right)^2, \quad j=1, 2, \dots, m,$$

while the within-imputation variance is denoted by $\bar{U}$ and is

$$\bar{U} = m^{-1} \sum U^{(j)}, \quad j=1, 2, \dots, m,$$

where $U(\hat{Q})$ is the estimated variance of $\hat{Q}^{(j)}$.

The total variance, denoted by T is then equal to

$$T = \bar{U} + (1 + m^{-1}) B. \quad (1)$$

In equation (1), the $(1 + m^{-1})B$ estimates the increase in variance due to the missing data and $\bar{U}$ estimates the variance if the data were complete [16].

In an ideal situation, where there could be an infinite number of imputations,

$$(\bar{Q} - Q) / \sqrt{T} \quad (2)$$

would have a $N(0, 1)$ distribution. However, to account for the finite number of imputations and for the simulation error associated with that finite number of repetitions, equation (2) follows a t_ν distribution. The degrees of freedom associated with this distribution is found by matching the first two moments of a chi-squared distribution to yield

$$\nu = (m - 1) \left(1 + \frac{\bar{U}}{(1 + m^{-1}) B} \right)^2 \quad (3)$$

[18, 22]. Adjustments to the degrees of freedom presented in equation (3) are presented by Barnard and Rubin [2], Lipsitz et al. [12], and Reiter [15] and are compared in Wagsta and Harel [25].

2.2. Two-Stage Multiple Imputation

Two-stage multiple imputation (or nested multiple imputation) involves generating imputations through a two step process to account for two different types of missing values. That is, consider a situation where the missing data are of two different types. These two types could be dropout in a longitudinal study and intermittent missed measurements. Another example is planned and unplanned values in survey studies. A third example could be missing responses and missing covariates. If the missing data are of two different types, it may be beneficial to treat them differently and adjust the multiple imputation method accordingly.

The general idea behind two-stage multiple imputation is to impute one type of missing values m times. Then, for each of those m data sets, impute the second type of missing values n times, treating the imputations for the first type of missing values as fixed and known. This yields a total of mn complete data sets from which estimates and variances are obtained [7].

Two types of missing values do not necessarily need to be treated separately. However, there are some advantages to this two-stage structure. One advantage is the ability to account for the degree of uncertainty contributed by each type of missing value. The second advantage is the ability to make different probabilistic assumptions on each of the types of missing values. This second advantage is particularly important because it allows the imputer to differentiate between missing values strongly related to the outcome being measured and values that are missing for some other reason irrelevant to the outcome. Furthermore, the estimates and standard errors should be the same whether one uses conventional multiple imputation or two-stage imputation (with the exception of simulation error) but what is gained is information about the distribution of the missing information because the missing data assumptions can be separated and rates of missing information due to each type of missing value can be estimated. The drawback of this approach is an increase in the required computation time.

Suppose the missing data are partitioned into two parts so that $Y = (Y_{obs}, Y_{mis}^A, Y_{mis}^B)$. Define M^+ as a random variable which is a matrix of missing data indicators of the same size as Y with 0 in each position corresponding to Y_{obs} , 1 in each position corresponding to Y_{mis}^A and 2 in each position corresponding to Y_{mis}^B . Then, to carry out the two-stage multiple imputation, first draw m values of Y_{mis}^A from

$$Y_{mis}^{A(j)} \sim P(Y_{mis}^A | Y_{obs}, M^+) \quad (4)$$

and then for each $Y_{mis}^{A(j)}$ draw n values of Y_{mis}^B from

$$Y_{mis}^{B(j,k)} \sim P(Y_{mis}^B | Y_{obs}, Y_{mis}^{A(j)}, M^+). \quad (5)$$

Under certain ignorability conditions [7, 9], Equations (4) and (5) can be reduced to $P(Y_{mis}^A | Y_{obs})$ and $P(Y_{mis}^B | Y_{obs}, Y_{mis}^A)$, respectively.

Shen [24] derived combining rules for this nested imputation scheme where he used the context of one set of imputations being computationally expensive and the other as being computationally inexpensive. Regardless of the application, the combining rules are reminiscent of a classical nested analysis of variance with Y_{mis}^A as a blocking factor [24].

The overall point estimate is

$$\bar{Q} = \frac{1}{mn} \sum_{j=1}^m \sum_{k=1}^n \hat{Q}^{(j,k)} = \frac{1}{m} \sum_{j=1}^m \bar{Q}_j.$$

where $\bar{Q}_j$ is the average of the m^{th} nest, which is similar in concept to the combining rule from standard multiple imputation. The variance of $\bar{Q}$, however, now has three components. There is the estimated complete data variance

$$\bar{U} = \frac{1}{mn} \sum_{j=1}^m \sum_{k=1}^n U^{(j,k)},$$

the between-block imputation variance

$$B = \frac{1}{m-1} \sum_{j=1}^m \left(\bar{Q}_{j\cdot} - \bar{Q} \right)^2,$$

and the within-block imputation variance

$$W = \frac{1}{m} \sum_{j=1}^m \frac{1}{n-1} \sum_{k=1}^n \left(\hat{Q}^{(j,k)} - \bar{Q}_{j\cdot} \right)^2.$$

Then, Shen [24] derived that the total variance is

$$T = \bar{U} + \left(1 + m^{-1}\right) B + \left(1 - n^{-1}\right) W$$

and from there could determine that inferences for Q should be based on a t -distribution;

$(Q - \bar{Q}) / \sqrt{T} \sim t_{\nu_*}$ with degrees of freedom

$$\nu_*^{-1} = \frac{1}{m-1} \left(\frac{(1+1/m) B}{T} \right)^2 + \frac{1}{m(n-1)} \left(\frac{(1-1/n) W}{T} \right)^2.$$

3. Methodological Derivation of Three-Stage Multiple Imputation

In this section, we describe the methodology for implementating three-stage multiple imputation and provide a proof of the methodology with a derivation of its corresponding limiting distribution.

3.1. Implementation of Three-Stage Multiple Imputation

The approach used for implementing multiple imputation in three stages begins with partitioning the data. Let Y_{com} represent the complete data set. This set can be partitioned into four parts: $(Y_{obs}, Y_{mis}^A, Y_{mis}^B, Y_{mis}^C)$ where Y_{obs} represents the observed part of the data, Y_{mis}^A represents the part of the data missing due to the first type of missingness, Y_{mis}^B represents the part of the data missing due to the second type of missingness and Y_{mis}^C represents the part of the data missing due to the third type of missingness. Let $M^\dagger$ represent the set of extended missingness indicators. $M^\dagger$ is an array the same size as Y_{com} which

contains a 0 in every position corresponding to an observed value, a 1 in every position corresponding to a value in Y_{mis}^A , a 2 in every position corresponding to a value in Y_{mis}^B , and a 3 in every position corresponding to a value in Y_{mis}^C . The joint model is now expressed in terms of M^+ and Y_{com} :

$$P(Y_{com}, M^+, \theta, \phi^+) = P(Y_{com} | \theta) P(M^+ | Y_{com}, \phi^+) P(\theta, \phi^+) \quad (6)$$

where θ represents the parameter of interest and ϕ^+ represents the parameters of the extended missingness mechanism.

To implement three-stage multiple imputation, we use an extension of the procedure used in two-stage multiple imputation [24]. First we draw L independent values of Y_{mis}^A from its predictive distribution,

$$Y_{mis}^{A(l)} \sim P(Y_{mis}^A | Y_{obs}, M^+), \quad l=1, 2, \dots, L. \quad (7)$$

Then, for each imputation of $Y_{mis}^{A(l)}$, we draw M conditionally independent values of Y_{mis}^B ;

$$Y_{mis}^{B(l,m)} \sim P(Y_{mis}^B | Y_{obs}, Y_{mis}^{A(l)}, M^+), \quad m=1, 2, \dots, M. \quad (8)$$

Finally, for each imputation of $(Y_{mis}^{A(l)}, Y_{mis}^{B(l,m)})$ we draw N conditionally independent values of Y_{mis}^C ;

$$Y_{mis}^{C(l,m,n)} \sim P(Y_{mis}^C | Y_{obs}, Y_{mis}^{A(l)}, Y_{mis}^{B(l,m)}, M^+), \quad n=1, 2, \dots, N \quad (9)$$

for a grand total of $L \times M \times N$ completed data sets.

Considered another way, what is being done is drawing L conventional imputations of Y_{mis} then, for each of those (holding one set fixed), generating $(M-1)$ additional draws of Y_{mis}^B , treating Y_{mis}^A as fixed and finally, for each of those (holding two sets fixed), generating $(N-2)$ additional draws of Y_{mis}^C , treating Y_{mis}^A and Y_{mis}^B as fixed. Under certain conditions described further in a 2013 University of Connecticut dissertation by the first author for three-stage multiple imputation and Harel and Schafer [9] for two-stage multiple imputation, the information in M^+ can be ignored fully or partially in one or more stages.

From a Bayesian point of view, the objective is to obtain the distribution of Q given the observed data, Y_{obs} , and M^+ . Let $Y_{com}^{(l,m,n)} = \{Y_{obs}, Y_{mis}^{A(l)}, Y_{mis}^{B(l,m)}, Y_{mis}^{C(l,m,n)}\}$ be the completed data sets. The posterior distribution of Q can be expressed as

$$\begin{aligned}
P(Q|Y_{obs}, M^+) &= \int P(Q|Y_{obs}, Y_{mis}, M^+) P(Y_{mis}|Y_{obs}, M^+) dY_{mis} \\
&= \lim_{L \rightarrow \infty} \frac{1}{L} \sum_{l=1}^L P(Q|Y_{obs}, Y_{mis}^{A(l)}, Y_{mis}^{B(l,1)}, Y_{mis}^{C(l,1,1)}, M^+) \\
&= \lim_{L \rightarrow \infty} \frac{1}{L} \sum_{l=1}^L P(Q|Y_{com}^{(l,1,1)}, M^+).
\end{aligned}$$

As with standard multiple imputation and two-stage multiple imputation, the goal is to use the completed data statistics to make inferences. Let

$$S_{LMN} = \left\{ \left(\hat{Q}^{(l,m,n)}, U^{(l,m,n)} \right) : l=1, \dots, L; m=1, \dots, M; n=1, \dots, N \right\}$$

denote the set of completed data statistics.

Since the number of imputations, L , is finite, there is a loss of information using S_{LMN} rather than Y_{obs} . The conditional distribution of Q given S_{LMN} and M^+ is

$$\begin{aligned}
P(Q|S_{LMN}, M^+) &= \int P(Q|S_{\infty}, S_{LMN}, M^+) P(S_{\infty}|S_{LMN}, M^+) dS_{\infty} \\
&= \int P(Q|S_{\infty}, M^+) P(S_{\infty}|S_{LMN}, M^+) dS_{\infty}
\end{aligned}$$

where S_{∞} is S_{LMN} when $L = \infty$.

The conditional distribution of Q given S_{LMN} and M^+ is difficult to derive analytically so it is approximated in five steps.

1. Derive $P(Q|S_{\infty}, M^+)$. (Section 3.2)
2. Derive $P(\bar{Q}_{\infty}, \bar{U}_{\infty}|S_{LMN}, B_{\infty}^{(b)}, B_{\infty}^{(w1)}, B_{\infty}^{(w2)}, M^+)$ where these variables are functions of S_{∞} to be defined in the derivation. (Section 3.3)
3. Combine results of the first two steps to get $P(Q|S_{LMN}, B_{\infty}^{(b)}, B_{\infty}^{(w1)}, B_{\infty}^{(w2)}, M^+)$. (Section 3.4)
4. Approximate $P(\bar{U}_{LMN} + (1 + \frac{1}{L}) B_{\infty}^{(b)} + (1 + \frac{1}{LM}) B_{\infty}^{(w1)} + (1 + \frac{1}{LMN}) B_{\infty}^{(w2)}|S_{LMN}, M^+)$. (Section 3.5)
5. Combine results of steps 3 and 4 to obtain an approximation to $P(Q|S_{LMN}, M^+)$. (Section 3.6)

The end goal is to approximate the posterior distribution of Q given the observed data.

3.2. Step 1

The first step of the derivation is to find the conditional distribution of Q given S_{∞} , M^+ . Conditional on the completed data (Y_{obs} , Y_{mis}), the completed data posterior distribution of

Q has mean $\hat{Q} = \hat{Q}(Y_{obs}, Y_{mis})$ and variance $U = U(Y_{obs}, Y_{mis})$. The actual posterior mean of Q is

$$E(Q|Y_{obs}, M^+) = E(\hat{Q}|Y_{obs}, M^+) \quad (10)$$

and the posterior variance of Q is

$$Var(Q|Y_{obs}, M^+) = E(U|Y_{obs}, M^+) + Var(\hat{Q}|Y_{obs}, M^+) \quad (11)$$

by the rules of iterative conditional expectation.

Suppose we have $L \times M \times N$ imputations where L is close to infinity. Then, we have $L \times M \times N$ sets of completed data statistics S_{LMN} as defined previously. The posterior mean of $\hat{Q}$ and the posterior mean of U are simulated in a manner similar to standard multiple imputation.

$$\bar{Q}_\infty \equiv \lim_{L \rightarrow \infty} \frac{1}{LMN} \sum_{l=1}^L \sum_{m=1}^M \sum_{n=1}^N \hat{Q}^{(l,m,n)} = E(\hat{Q}|Y_{obs}, M^+) \quad (12)$$

$$\bar{U}_\infty \equiv \lim_{L \rightarrow \infty} \frac{1}{LMN} \sum_{l=1}^L \sum_{m=1}^M \sum_{n=1}^N \hat{U}^{(l,m,n)} = E(\hat{U}|Y_{obs}, M^+). \quad (13)$$

Since the $L \times M \times N$ imputations are not independent, we must be cautious about simulating $Var(\hat{Q}|Y_{obs}, M^+)$. Using Barnard [1] orthogonal decomposition, $\hat{Q}$ can be decomposed into four orthogonal pieces.

$$\begin{aligned} \hat{Q}(Y_{obs}, Y_{mis}^{A(l)}, Y_{mis}^{B(l,m)}, Y_{mis}^{C(l,m,n)}, M^+) \\ = H_0(Y_{obs}, M^+) \\ + H_1(Y_{mis}^{A(l)}, Y_{obs}, M^+) \\ + H_2(Y_{mis}^{A(l)}, Y_{mis}^{B(l,m)}, Y_{obs}, M^+) \\ + H_3(Y_{mis}^{A(l)}, Y_{mis}^{B(l,m)}, Y_{mis}^{C(l,m,n)}, Y_{obs}, M^+) \end{aligned} \quad (14)$$

where

$$\begin{aligned}
H_0(Y_{obs}, M^+) &= E(\hat{Q}|Y_{obs}, M^+) \\
H_1(Y_{obs}, Y_{mis}^{A(l)}, M^+) &= E(\hat{Q}|Y_{obs}, Y_{mis}^{A(l)}, M^+) - H_0 \\
H_2(Y_{obs}, Y_{mis}^{A(l)}, Y_{mis}^{B(l,m)}, M^+) &= E(\hat{Q}|Y_{obs}, Y_{mis}^{A(l)}, Y_{mis}^{B(l,m)}, M^+) - H_1 \\
H_3(Y_{obs}, Y_{mis}^{A(l)}, Y_{mis}^{B(l,m)}, Y_{mis}^{C(l,m,n)}, M^+) &= E(\hat{Q}|Y_{obs}, Y_{mis}^{A(l)}, Y_{mis}^{B(l,m)}, Y_{mis}^{C(l,m,n)}, M^+) - H_2.
\end{aligned}$$

Using this construction, all of the pairwise correlations are zero. $Var(\hat{Q}|Y_{obs}, M^+)$ can be expressed as

$$Var(\hat{Q}|Y_{obs}, M^+) = Var(H_1|Y_{obs}, M^+) + Var(H_2|Y_{obs}, M^+) + Var(H_3|Y_{obs}, M^+).$$

(15)

Using the analysis of variance results (Table 1) for three-stage nested design,

$$\begin{aligned}
Var(H_3|Y_{obs}, M^+) &= B_{\infty}^{(w2)} \equiv \lim_{L \rightarrow \infty} MS_{N(M)} \\
Var(H_2|Y_{obs}, M^+) &= B_{\infty}^{(w1)} \equiv \lim_{L \rightarrow \infty} \frac{MS_{M(L)} - MS_{N(M)}}{N} \\
Var(H_1|Y_{obs}, M^+) &= B_{\infty}^{(b)} \equiv \lim_{L \rightarrow \infty} \frac{MS_L - MS_{M(L)}}{MN}.
\end{aligned}$$

Therefore, the posterior variance of $\hat{Q}$ is

$$Var(\hat{Q}|Y_{obs}, M^+) = B_{\infty}^{(b)} + B_{\infty}^{(w1)} + B_{\infty}^{(w2)} = B_{\infty}. \quad (16)$$

From equations (10) and (12), we conclude that

$$E(Q|Y_{obs}, M^+) = \bar{Q}_{\infty} \quad (17)$$

and from equations (11), (12), and (16),

$$Var(Q|Y_{obs}, M^+) = \bar{U}_{\infty} + B_{\infty} = T_{\infty}. \quad (18)$$

Then, assuming an approximate normal posterior distribution, we have

$$(Q|Y_{obs}, M^+) \sim N(\bar{Q}_{\infty}, T_{\infty}). \quad (19)$$

Under the normal posterior distribution, this posterior distribution is equivalent to the

conditional distribution of Q given $h(S_\infty) = \left(\bar{Q}_\infty, \bar{U}_\infty, B_\infty^{(b)}, B_\infty^{(w1)}, B_\infty^{(w2)}\right)$ which is what we wanted to show.

3.3. Step 2

The second step of the derivation is to find the distribution of

$$\left(\bar{Q}_\infty, \bar{U}_\infty\right) | \left(S_{LMN}, B_\infty^{(b)}, B_\infty^{(w1)}, B_\infty^{(w2)}, M^+\right). \quad (20)$$

In order to do this, we must first find the conditional distribution of S_{LMN} given

$\left(\bar{Q}_\infty, \bar{U}_\infty, B_\infty^{(b)}, B_\infty^{(w1)}, B_\infty^{(w2)}\right)$ and then treat $(\bar{Q}_\infty, \bar{U}_\infty)$ as estimands and S_{LMN} as data. Then, by assuming a flat prior on $\bar{Q}_\infty$, we can obtain the posterior distribution of $(\bar{Q}_\infty, \bar{U}_\infty)$.

From the orthogonal decomposition (14), it can be shown that the covariance between any two $\hat{Q}$ s is as follows:

$$Cov\left(\hat{Q}^{(l_1, m_1, n_1)}, \hat{Q}^{(l_2, m_2, n_2)} | Y_{obs}, M^+\right) = \begin{cases} 0: & l_1 \neq l_2 \\ B_\infty^{(b)}: & l_1 = l_2, m_1 \neq m_2 \\ B_\infty^{(b)} + B_\infty^{(w1)}: & l_1 = l_2, m_1 = m_2, n_1 \neq n_2 \\ B_\infty^{(b)} + B_\infty^{(w1)} + B_\infty^{(w2)}: & l_1 = l_2, m_1 = m_2, n_1 = n_2. \end{cases}$$

Let $\hat{Q}^{(l)}$ denote all the $\hat{Q}$ s in the l^{th} nest. That is,

$$\hat{Q}^{(l)} = \left(\hat{Q}^{(l,1,1)}, \hat{Q}^{(l,1,2)}, \dots, \hat{Q}^{(l,M,N)}\right)^T.$$

The variance-covariance matrix of $\hat{Q}^{(l)}$ conditioned on $\bar{Q}_\infty, \bar{U}_\infty, B_\infty^{(b)}, B_\infty^{(w1)}$, and $B_\infty^{(w2)}$ can be expressed as

$$Var\left(\hat{Q}^{(l)} | Y_{obs}, M^+\right) = B_\infty^{(b)} 1_{MN \times MN} + \bigoplus_{i=1}^M \left[B_\infty^{(w1)} 1_{N \times N} + B_\infty^{(w2)} I_{N \times N} \right] \quad (21)$$

where $1_{MN \times MN}$ and $1_{N \times N}$ are square matrices with all the elements 1 and $I_{N \times N}$ represents the direct sum of the matrices which creates matrix with dimension $N \times N$. The $\bigoplus_{i=1}^M$ represents the direct sum of the matrices which creates a block diagonal matrix.

The conditional expectation of each element in $\hat{Q}^{(l)}$ is

$$E\left(\hat{Q}^{(l,m,n)} | Y_{obs}, M^+\right) = \bar{Q}_\infty. \quad (22)$$

Note that vectors from different nests are independent. That is, $\hat{Q}^{(l)}$ is independent of $\hat{Q}^{(h)}$. With the assumption of large-sample data, we can assume that given (Y_{obs}, M^+) , $\hat{Q}^{(l)}$ are independent and identically distributed draws from

$$\left(\hat{Q}^{(l)} | \bar{Q}_\infty, B_\infty^{(b)}, B_\infty^{(w1)}, B_\infty^{(w2)}\right) \sim N\left(\bar{Q}_\infty 1_{MN}, B_\infty^{(b)} 1_{MN \times MN} + \bigoplus_{i=1}^M \left[B_\infty^{(w1)} 1_{N \times N} + B_\infty^{(w2)} I_{N \times N}\right]\right). \quad (23)$$

Then the conditional expectation and variance follow:

$$\begin{aligned} E\left(\bar{Q}_{LMN} | Y_{obs}, M^+\right) &= E\left(\frac{1}{LMN} \sum_{l=1}^L \sum_{m=1}^M \sum_{n=1}^N \hat{Q}^{(l,m,n)}\right) \\ &= \bar{Q}_\infty \end{aligned} \quad (24)$$

and

$$\begin{aligned} Var\left(\bar{Q}_{LMN} | Y_{obs}, M^+\right) &= Var\left(\frac{1}{LMN} \sum_{l=1}^L \sum_{m=1}^M \sum_{n=1}^N \hat{Q}^{(l,m,n)}\right) \\ &= \frac{1}{L} B_\infty^{(b)} + \frac{1}{LM} B_\infty^{(w1)} + \frac{1}{LMN} B_\infty^{(w2)}. \end{aligned} \quad (25)$$

Since $\bar{Q}_{LMN}$ is a linear combination of jointly normal random variables, its sampling distribution also follows a normal distribution of the form

$$\left(\bar{Q}_{LMN} | \bar{Q}_\infty, B_\infty^{(b)}, B_\infty^{(w1)}, B_\infty^{(w2)}, M^+\right) \sim N\left(\bar{Q}_\infty, \frac{1}{L} B_\infty^{(b)} + \frac{1}{LM} B_\infty^{(w1)} + \frac{1}{LMN} B_\infty^{(w2)}\right). \quad (26)$$

The joint distribution of $U^{(l,m,n)}$ is much harder to specify but it is not necessary to do so. Instead, we use the assumption of lower order of variability in the same way as standard multiple imputation:

$$\left(U^{(l,m,n)} | Y_{obs}, M^+\right) \sim \left(\bar{U}_\infty, \ll B_\infty\right) \quad (27)$$

where $D \sim (A, \ll C)$ means that the distribution of D is centered around A with each component having variability substantially less than each positive component of C .

Now, we treat $(\bar{Q}_\infty, \infty)$ as two unknown estimands with a conditional distribution that can be found by

$$P\left(\bar{Q}_\infty, \bar{U}_\infty | S_{LMN}, B_\infty^{(b)}, B_\infty^{(w1)}, B_\infty^{(w2)}, M^+\right) \propto P\left(\bar{Q}_\infty, \bar{U}_\infty\right) P\left(S_{LMN} | \bar{Q}_\infty, \bar{U}_\infty, B_\infty^{(b)}, B_\infty^{(w1)}, B_\infty^{(w2)}, M^+\right).$$

(28)

Assume a flat prior for $\bar{Q}_\infty$ and combine with the sampling distribution of $\bar{Q}_{LMN}$ to get

$$\left(\bar{Q}_\infty | S_{LMN}, B_\infty^{(b)}, B_\infty^{(w1)}, B_\infty^{(w2)}, M^+\right) \sim N\left(\bar{Q}_{LMN}, \frac{1}{L} B_\infty^{(b)} + \frac{1}{LM} B_\infty^{(w1)} + \frac{1}{LMN} B_\infty^{(w2)}\right).$$

(29)

Also, assuming a relatively diffuse prior on $\bar{U}_\infty$ combined with the sampling distribution of $U^{(l,m,n)}$, we get

$$\left(\bar{U}_\infty | S_{LMN}, B_\infty^{(b)}, B_\infty^{(w1)}, B_\infty^{(w2)}, M^+\right) \sim \left(\bar{U}_{LMN}, \ll \frac{B_\infty}{LMN}\right). \quad (30)$$

3.4. Step 3

In this step of the derivation, the results from the previous two steps are combined to get the conditional distribution of Q given $(S_{LMN}, B_\infty^{(b)}, B_\infty^{(w1)}, B_\infty^{(w2)}, M^+)$. This is achieved with the following integration:

$$P\left(Q | S_{LMN}, B_\infty^{(b)}, B_\infty^{(w1)}, B_\infty^{(w2)}, M^+\right) \quad (31)$$

$$= \int P\left(Q | \bar{Q}_\infty, \bar{U}_\infty, B_\infty^{(b)}, B_\infty^{(w1)}, B_\infty^{(w2)}, S_{LMN}, M^+\right) \times P\left(\bar{Q}_\infty, \bar{U}_\infty | B_\infty^{(b)}, B_\infty^{(w1)}, B_\infty^{(w2)}, S_{LMN}, M^+\right) d\bar{Q}_\infty d\bar{U}_\infty$$

$$= \int P\left(Q | \bar{Q}_\infty, \bar{U}_\infty, B_\infty^{(b)}, B_\infty^{(w1)}, B_\infty^{(w2)}, M^+\right) \times P\left(\bar{Q}_\infty, \bar{U}_\infty | B_\infty^{(b)}, B_\infty^{(w1)}, B_\infty^{(w2)}, S_{LMN}, M^+\right) d\bar{Q}_\infty d\bar{U}_\infty.$$

(32)

Note that the distribution of $\left(Q | \bar{Q}_\infty, \bar{U}_\infty, B_\infty^{(b)}, B_\infty^{(w1)}, B_\infty^{(w2)}, S_{LMN}, M^+\right)$ is $N(\bar{Q}_\infty, \bar{U}_\infty + B_\infty)$ because of the lower variability result. Combined with the conditional distribution of $\bar{Q}_\infty$, we get

$$(Q|S_{LMN}, B_{\infty}^{(b)}, B_{\infty}^{(w1)}, B_{\infty}^{(w2)}, M^+) \sim N\left(\bar{Q}_{LMN}, \bar{U}_{LMN} + \left(1 + \frac{1}{L}\right) B_{\infty}^{(b)} + \left(1 + \frac{1}{LM}\right) B_{\infty}^{(w1)} + \left(1 + \frac{1}{LMN}\right) B_{\infty}^{(w2)}\right).$$

(33)

3.5. Step 4

The plan for this step of the derivation is to approximate the variance, which is a function of $B_{\infty}^{(b)}$, $B_{\infty}^{(w1)}$, and $B_{\infty}^{(w2)}$, by an inverse χ^2 distribution.

Define the expected mean squares as

$$\begin{aligned} ES^{(b)} &= E(MS_L) \\ ES^{(w1)} &= E(MS_{M(L)}) \\ ES^{(w2)} &= E(MS_{N(M)}). \end{aligned} \quad (34)$$

Assuming the improper prior

$$\pi(ES^{(b)}, ES^{(w1)}, ES^{(w2)}) \propto \frac{1}{ES^{(b)}} \times \frac{1}{ES^{(w1)}} \times \frac{1}{ES^{(w2)}} \quad (35)$$

and the analysis of variance table, we have

$$\begin{aligned} ES^{(b)} &\sim \chi^{-2}(L-1, MS_L) \\ ES^{(w1)} &\sim \chi^{-2}(L(M-1), MS_{M(L)}) \\ ES^{(w2)} &\sim \chi^{-2}(LM(N-1), MS_{N(M)}) \end{aligned} \quad (36)$$

where $\chi^{-2}(\nu, s)$ denotes a scaled inverse χ^2 distribution with degree of freedom ν and scale s .

Also from the analysis of variance table, we have

$$\begin{aligned} ES^{(b)} &= B_{\infty}^{(w2)} + NB_{\infty}^{(w1)} + MNB_{\infty}^{(b)} \\ ES^{(w1)} &= B_{\infty}^{(w2)} + NB_{\infty}^{(w1)} \\ ES^{(w2)} &= B_{\infty}^{(w2)}. \end{aligned} \quad (37)$$

Now the conditional variance of Q given $(S_{LMN}, B_{\infty}^{(b)}, B_{\infty}^{(w1)}, B_{\infty}^{(w2)}, M^+)$ can be rewritten as

$$\frac{1}{MN} \left(1 + \frac{1}{L}\right) ES^{(b)} + \frac{1}{N} \left(1 - \frac{1}{M}\right) ES^{(w1)} + \left(1 - \frac{1}{N}\right) ES^{(w2)}. \quad (38)$$

Note that if $N = 1$, the above equation reduces to the form for two-stage multiple imputation and if $N = M = 1$, the above equation reduces to the form for standard multiple imputation.

The goal is to rewrite the conditional variance of Q as a linear combination of three scaled inverse χ^2 's plus a constant and then integrate out $ES^{(b)}$, $ES^{(w1)}$, and $ES^{(w2)}$ over the conditional distribution of $(ES^{(b)}, ES^{(w1)}, ES^{(w2)}) | S_{LMN}, M^+$. However, this is challenging given the constraint $ES^{(b)} - ES^{(w1)} - ES^{(w2)}$. Instead, the approach from standard multiple imputation is used and we approximate the conditional variance by a scaled inverse χ^2 . By relaxing the constraint and assuming that $ES^{(b)}$, $ES^{(w1)}$, and $ES^{(w2)}$ are independent scaled inverse χ^2 's, we can match the first two moments. From the analysis of variance table, we have

$$\begin{aligned} MS_L &\sim ES^{(b)} \chi^2 [L - 1] / (L - 1) \\ MS_{M(L)} &\sim ES^{(w1)} \chi^2 [L(M - 1)] / (L(M - 1)) \\ MS_{N(M)} &\sim ES^{(w2)} \chi^2 [LM(N - 1)] / (LM(N - 1)). \end{aligned} \quad (39)$$

For convenience, a conjugate prior distribution is chosen for $ES^{(b)}$, $ES^{(w1)}$, $ES^{(w2)}$ and has the form:

$$\begin{aligned} \pi \left(ES^{(b)}, ES^{(w1)}, ES^{(w2)} | r^{(b)}, r^{(w1)}, r^{(w2)}, \omega^{(b)}, \omega^{(w1)}, \omega^{(w2)} \right) &\propto \left(ES^{(b)} \right)^{-\left(\frac{w^{(b)}}{2} + 1 \right)} \exp \left[-\frac{r^{(b)} \omega^{(b)}}{2 ES^{(b)}} \right] \\ &\times \left(ES^{(w1)} \right)^{-\left(\frac{w^{(w1)}}{2} + 1 \right)} \exp \left[-\frac{r^{(w1)} \omega^{(w1)}}{2 ES^{(w1)}} \right] \\ &\times \left(ES^{(w2)} \right)^{-\left(\frac{w^{(w2)}}{2} + 1 \right)} \exp \left[-\frac{r^{(w2)} \omega^{(w2)}}{2 ES^{(w2)}} \right] \end{aligned} \quad (40)$$

where $r^{(b)}$, $\omega^{(b)}$, $r^{(w1)}$, $\omega^{(w1)}$, $r^{(w2)}$, and $ES^{(w2)}$, respectively.

Then, combining equations (39) and (40), we get

$$\begin{aligned}
\pi \left(ES^{(b)}, ES^{(w1)}, ES^{(w2)} | S_{LMN} \right) &\propto \left(ES^{(b)} \right)^{-\left(\frac{(L-1)+\omega^{(b)}}{2} + 1 \right)} \exp \left[-\frac{(L-1) MS_L + r^{(b)} \omega^{(b)}}{2 ES^{(b)}} \right] \\
&\times \left(ES^{(w1)} \right)^{-\left(\frac{L(M-1)+\omega^{(w1)}}{2} + 1 \right)} \exp \left[-\frac{L(M-1) MS_{M(L)} + r^{(w1)} \omega^{(w1)}}{2 ES^{(w1)}} \right] \\
&\times \left(ES^{(w2)} \right)^{-\left(\frac{LM(N-1)+\omega^{(w2)}}{2} + 1 \right)} \exp \left[-\frac{LM(N-1) MS_{N(M)} + r^{(w2)} \omega^{(w2)}}{2 ES^{(w2)}} \right]
\end{aligned}
\tag{41}$$

where $ES^{(b)}, ES^{(w1)}, ES^{(w2)}$. Without that constraint, $ES^{(b)}, ES^{(w1)}$, and $ES^{(w2)}$ will be independently distributed as scaled inverse χ^2 distributions. The detailed derivation is found in Appendix A. From that derivation, we use the approximation

$$\bar{U}_{LMN} + \left(1 + \frac{1}{L} \right) B_{\infty}^{(b)} + \left(1 + \frac{1}{LM} \right) B_{\infty}^{(w1)} + \left(1 + \frac{1}{LMN} \right) B_{\infty}^{(w2)} | S_{LMN} \sim \chi^{-2}(\nu, T_{LMN}).
\tag{42}$$

Combined with the improper prior (35) and without the constraint that $ES^{(b)}, ES^{(w1)}, ES^{(w2)}$, the inverse χ^2 distribution has scale parameter

$$T_{LMN} = \bar{U}_{LMN} + \frac{1}{MN} \left(1 + \frac{1}{L} \right) MS_L + \frac{1}{N} \left(1 - \frac{1}{M} \right) MS_{M(L)} + \left(1 - \frac{1}{N} \right) MS_{N(M)} \tag{43}$$

and degrees of freedom

$$\begin{aligned}
\nu^{-1} = & \left[\frac{\frac{1}{MN} \left(1 + \frac{1}{L} \right) MS_L}{T_{LMN}} \right]^2 (L-1)^{-1} + \left[\frac{\frac{1}{N} \left(1 - \frac{1}{M} \right) MS_{M(L)}}{T_{LMN}} \right]^2 (L(M-1))^{-1} \\
& + \left[\frac{\left(1 - \frac{1}{N} \right) MS_{N(M)}}{T_{LMN}} \right]^2 (LM(N-1))^{-1}.
\end{aligned}
\tag{44}$$

Note that if $N=1$ this reduces to the two-stage case and if $N=M=1$ this reduces to standard multiple imputation.

3.6. Step 5

The final step is to arrive at the conditional distribution of Q given S_{LMN} , by integrating the conditional distribution of

$$Q|\bar{U}_{LMN} + \left(1 + \frac{1}{L}\right) B_{\infty}^{(b)} + \left(1 + \frac{1}{LM}\right) B_{\infty}^{(w1)} + \left(1 + \frac{1}{LMN} B_{\infty}^{(w2)}\right), S_{LMN}, M^+$$

over the conditional distribution in

$$\bar{U}_{LMN} + \left(1 + \frac{1}{L}\right) B_{\infty}^{(b)} + \left(1 + \frac{1}{LM}\right) B_{\infty}^{(w1)} + \left(1 + \frac{1}{LMN} B_{\infty}^{(w2)}\right) | S_{LMN}, M^+$$

to get

$$(Q|S_{LMN}, M^+) \sim t_{\nu}(\bar{Q}_{LMN}, T_{LMN}). \quad (45)$$

The $100(1 - \alpha)\%$ interval estimate for Q is

$$\left(\bar{Q}_{LMN} - t_{\nu}(100(1 - \alpha/2)) \sqrt{T_{LMN}}, \bar{Q}_{LMN} + t_{\nu}(100(1 - \alpha/2)) \sqrt{T_{LMN}}\right). \quad (46)$$

3.7. Alternative Notation

For consistency with the notation of [18] and [7], the results can be expressed as follows:

$$\begin{aligned} \bar{Q} &= \frac{1}{LMN} \sum_{l=1}^L \sum_{m=1}^M \sum_{n=1}^N \hat{Q}^{(l,m,n)} \\ \bar{U} &= \frac{1}{LMN} \sum_{l=1}^L \sum_{m=1}^M \sum_{n=1}^N U^{(l,m,n)} \\ B &= \frac{1}{L-1} \sum_{l=1}^L \left(\bar{Q}_{l..} - \bar{Q}_{...}\right)^2 \\ W_1 &= \frac{1}{L(M-1)} \sum_{l=1}^L \sum_{m=1}^M \left(\bar{Q}_{lm.} - \bar{Q}_{l..}\right)^2 \\ W_2 &= \frac{1}{LM(N-1)} \sum_{l=1}^L \sum_{m=1}^M \sum_{n=1}^N \left(\hat{Q}^{(l,m,n)} - \bar{Q}_{lm.}\right)^2. \end{aligned}$$

T_{LMN} can be expressed as

$$T = \bar{U}_{LMN} + \left(1 + \frac{1}{L}\right) B + \left(1 - \frac{1}{M}\right) W_1 + \left(1 - \frac{1}{N}\right) W_2 \quad (47)$$

and the degrees of freedom can be expressed as

$$\nu^{-1} = \left[\frac{\left(1 + \frac{1}{L}\right) B}{T} \right]^2 (L-1)^{-1} + \left[\frac{\left(1 - \frac{1}{M}\right) W_1}{T} \right]^2 (L(-1))^{-1} + \left[\frac{\left(1 - \frac{1}{N}\right) W_2}{T} \right]^2 (LM(N-1))^{-1}. \quad (48)$$

4. Extension for k -Stage Multiple Imputation

The proof for the combining rules for k -stage multiple imputation is a straightforward extension of the proof for three-stage multiple imputation. The details of the proof are not included here but directly follow the steps presented in Section 3 with additional variables. The process for creating the imputations would be similar but would have additional nested steps for creating the completed data sets. A discussion of the benefits and complications of including k types of missing values is beyond the scope of this paper but the results for the combining rules are presented here.

Let $N_1, N_2, \dots, N_k$ denote the number of imputations at each of the stages of imputation for a total of $N_1 \times N_2 \times \dots \times N_k$ completed data sets. Let $\bar{Q}$ be the average of the parameter estimates from each of the completed data sets and $\bar{V}$ be the average of the complete data variance estimates. Let MS_i denote the mean square from the i^{th} nest in the nested ANOVA decomposition where $i = 1, \dots, k$. Then, the total variance, T , is

$$T = \bar{U} + \frac{1}{\prod_{j=2}^k N_j} \left(1 + \frac{1}{N_1}\right) MS_1 + \sum_{i=2}^{k-1} \frac{1}{\prod_{j=i+1}^k N_j} \left(1 - \frac{1}{N_i}\right) MS_i + \left(1 - \frac{1}{N_k}\right) MS_k \quad (49)$$

with degrees of freedom

$$\begin{aligned} \nu^{-1} = & \left[\frac{\frac{1}{\prod_{j=2}^k N_j} \left(1 + \frac{1}{N_1}\right) MS_1}{T} \right]^2 [(N_1 - 1)]^{-1} \\ & + \sum_{i=2}^{k-1} \left[\left[\frac{\frac{1}{\prod_{j=i+1}^k N_j} \left(1 - \frac{1}{N_i}\right) MS_i}{T} \right]^2 \left[\left(\prod_{j=1}^{k-2} N_j \right) (N_i - 1) \right]^{-1} \right] \\ & + \left[\frac{\left(1 - \frac{1}{N_k}\right) MS_k}{T} \right]^2 \left[\left(\prod_{j=1}^{k-1} N_j \right) (N_k - 1) \right]^{-1}. \end{aligned} \quad (50)$$

These formulae hold for three or more types of missing values. When $k = 3$, these formulae reduce to the three-stage multiple imputation combining rules in Equations (43) and (44). Similarly, when $k = 1$ or $k = 2$, these formulae reduce to the standard and two-stage combining rules previously discussed.

5. Ignorability

This section includes a brief overview of missing data terminology including *mechanisms of missingness*, *distinctness*, and *ignorability*. It also includes a three-stage extension of the definition of ignorability with proof.

5.1. Terminology Overview

The mechanism of missingness describes the relationship between the probability of a value being missing and the other variables in the data set.

Let Y represent the complete data that can be partitioned as (Y_{obs}, Y_{mis}) where Y_{obs} is the observed part of Y and Y_{mis} is the missing part of Y . Let R be an indicator random variable (or matrix) indicating whether or not Y is observed or missing. Let $R = 1$ denote a value which is observed and let $R = 0$ denote a value which is missing. The statistical model for missing data is $P(R | Y; \phi)$ where ϕ is the parameter for the missing data process. The mechanism of missingness is determined by the dependency of R on the variables in the data set. The following are the mechanisms of missingness as described in Rubin [17] and Rubin [18].

The first mechanism of missingness is *missing at random*. This mechanism of missingness is given by

$$P(R | Y; \phi) = P(R | Y_{obs}; \phi).$$

That is, the probability of missingness is only dependent on observed values in Y and not on any unobserved values in Y . A simple example of MAR is a survey where subjects over a certain age refuse to answer a particular survey question and age is an observed covariate.

The second mechanism of missingness is a special case of missing at random known as *missing completely at random*. In this case, the mechanism of missingness is given by

$$P(R | Y; \phi) = P(R; \phi).$$

The probability of missingness is not dependent on any observed or unobserved values in Y . It is what one colloquially thinks of as “random.” One example of missing completely at random might be a computer malfunction that arbitrarily deletes some of the data values.

The third mechanism of missingness is referred to as *missing not at random*. This mechanism of missingness occurs when the conditions of missing at random are violated so that the probability of missingness is dependent on Y_{mis} or some unobserved covariate. One instance of missing not at random might be subjects who have an income above a certain value refusing to report an income in the survey. Here the missingness is dependent on the unobserved response, income.

A missing data mechanism can also be classified as ignorable or nonignorable. In order for a missing data mechanism to be ignorable, two conditions should be met. The first condition is

that the data are missing at random or missing completely at random. The second condition is that the parameter of interest, θ , and the parameter of the missing data process, ϕ , are distinct. These parameters are considered distinct if the following condition holds:

$$P(\theta, \phi) = P(\theta) P(\phi),$$

that is, the joint prior distribution for θ and ϕ is equal to the product of two independent priors. From a frequentist perspective, the parameters are distinct if the joint parameter space is the Cartesian cross-product of the individual parameter spaces of θ and ϕ [21]. A missing data mechanism is classified as nonignorable if at least one of these conditions is not met. Ignorability represents the weakest set of conditions under which the distribution of R , the missingness, does not need to be considered in Bayesian or likelihood-based inference of θ [18].

5.2. Extended Ignorability

Rubin [17] defines missing at random (MAR) as follows:

Definition 1—Let M^+ be a set of indicator random variables that separate the complete data Y_{com} into (Y_{obs}, Y_{mis}) and let ϕ^+ be the parameter of the conditional distribution of M^+ given Y_{com} . Data are considered to be MAR if

$$P(M^+ | Y_{obs}, Y_{mis}, \phi^+) = P(M^+ | Y_{obs}, \phi^+) \quad (51)$$

for all possible values of ϕ^+ , at the realized values of M^+ and Y_{obs} .

Conceptually, this equates to the distribution of missingness being functionally independent of the missing part of the data. A missing data mechanism is said to be ignorable if MAR holds and θ and ϕ^+ are distinct [13]. Under ignorability, the model for M^+ given Y_{com} is irrelevant for likelihood-based or Bayesian inferences about θ . That is, the joint likelihood for (Y_{obs}, M^+) given (θ, ϕ^+) is

$$\begin{aligned} P(Y_{obs}, M^+ | \theta, \phi^+) &= \int P(Y_{com}, M^+ | \theta, \phi^+) dY_{mis} \\ &= \int P(Y_{obs}, Y_{mis} | \theta) P(M^+ | Y_{obs}, Y_{mis}, \phi^+) dY_{mis} \\ &= P(M^+ | Y_{obs}, \phi^+) \int P(Y_{obs}, Y_{mis} | \theta) dY_{mis} \\ &= P(M^+ | Y_{obs}, \phi^+) P(Y_{obs} | \theta) \\ &\propto L(\phi^+ | M^+, Y_{obs}) L(\theta | Y_{obs}) \end{aligned}$$

where $L(\phi^+ | M^+, Y_{obs})$ is a likelihood function that does not involve θ and $L(\theta | Y_{obs})$ is a likelihood function that does not involve ϕ^+ .

Harel [7] and Harel and Schafer [9] describe extended missingness when there are two types of missing values. Let the complete data be partitioned into $(Y_{obs}, Y_{mis}^A, Y_{mis}^B)$ and let M^+ be the extended missing data matrix. Let ϕ^+ be the parameter of the missing data process.

Definition 2—The missing data $Y_{mis} = (Y_{mis}^A, Y_{mis}^B)$ are said to be MAR^+ if

$$P(M^+ | Y_{obs}, Y_{mis}, \phi^+) = P(M^+ | Y_{obs}, \phi^+) \quad (52)$$

for all possible ϕ^+ , at the realized values of M^+ and Y_{obs} .

This condition is stronger than MAR because it assumes that the process that subdivides Y_{mis} does not depend on Y_{mis} either.

Harel [7] concludes that if θ and ϕ^+ are distinct and MAR^+ holds, then the information contained in M^+ can be ignored when making likelihood or Bayesian inferences for θ . The implication here is that M^+ may be ignored when imputing both Y_{mis}^A and Y_{mis}^B .

When three types of missing values are present, the data can be partitioned as

$Y_{com} = (Y_{obs}, Y_{mis}^A, Y_{mis}^B, Y_{mis}^C)$. M^+ remains the set of indicators partitioning the complete data. The purpose of describing extended missingness is not to encourage modeling of M^+ but instead to illustrate conditions under which the modeling of M^+ is unnecessary.

Definition 2 for the extended ignorability still holds with three types of missing values. The main idea of the definition is that the process that subdivides Y_{mis} is not dependent on Y_{mis} so the definition is relevant whether the missing data are being divided into two parts, three parts, or more. For completeness, Definition 2 can be rewritten as follows:

Definition 3—The missing data Y_{mis} , regardless of the number of partitions, are said to be MAR^+ if

$$P(M^+ | Y_{obs}, Y_{mis}, \phi^+) = P(M^+ | Y_{obs}, \phi^+)$$

for all possible ϕ^+ , at the realized values of M^+ and Y_{obs} .

Definition 2 is simply a special case of Definition 3 with two partitions and three-stage ignorability is a special case with three partitions.

Theorem 1—If θ and ϕ^+ are distinct and MAR^+ holds, then we can ignore the information contained in M^+ when making likelihood or Bayesian inferences about θ

The proof of this theorem is identical to that of Rubin's.

Theorem 1 shows that M^+ can be ignored at all imputation stages as indicated by the following corollary:

Corollary 1—If θ and ϕ^+ are distinct and MAR^+ holds, then we can ignore the information contained in M^+ in all stages of imputation (7), (8), (9), so that

$$P\left(Y_{mis}^A | Y_{obs}, M^+\right) = P\left(Y_{mis}^A | Y_{obs}\right)$$

and

$$P\left(Y_{mis}^B | Y_{obs}, Y_{mis}^A, M^+\right) = P\left(Y_{mis}^B | Y_{obs}, Y_{mis}^A\right)$$

and

$$P\left(Y_{mis}^C | Y_{obs}, Y_{mis}^A, Y_{mis}^B, M^+\right) = P\left(Y_{mis}^C | Y_{obs}, Y_{mis}^A, Y_{mis}^B\right).$$

The proof of this corollary can be found in Appendix B.

Weaker conditions under which M^+ can be ignored in one or more stages have been defined in Boyko [4] but are beyond the scope of this paper.

6. Simulations

6.1. Simulation Set-Up

In order to evaluate the effectiveness of the method proposed, simulations were run under varying scenarios to assess viability. Among the questions of interest are “Does the number of imputations significantly affect bias and coverage?” and “Does the order of imputation significantly affect bias and coverage?” All of the simulations in this section assume ignorability.

Data for the simulations are generated through the following steps. X_1 is generated from a $N(50, 100)$ and X_2 is generated from $N(20, 100)$. 100 values of X_1 and X_2 are randomly drawn. Y is created by taking $Y = 2X_1 + 3X_2$ so that Y has a known mean of 160. Then, Y_{com} is the data which includes Y and X_1 .

A missingness structure is imposed as follows:

1. The first type of missing value is created with a missing completely at random structure to simulate a missing covariate. That is, a prespecified percentage of the values in X_1 are randomly deleted. Let $MCAR\%$ denote the percentage of missing values due to the first type of missingness.
2. The second type of missing value is created under a missing at random structure. Let $MAR_1\%$ denote the percentage of missing values due to the first type of missing at random variable. Values in Y are removed if they are above the top $MAR_1\%$ percentile of X_1 .
3. The third type of missing value is created under a missing at random structure. Let $MAR_2\%$ denote the percentage of missing values due to the second type of missing at random variable. Values in Y are removed if they are below the bottom $MAR_2\%$ percentile of X_1 .

The missing values are imputed using the norm package in R [20] with varying numbers of imputations at each stage denoted by the ordered triple (L, M, N) and such that the order is $MCAR\%$, $MAR_1\%$, $MAR_2\%$.

The parameter of interest, Q , is the mean of Y . Data are generated, imputed, and analyzed 1000 times for percent bias, mean square error, and coverage. Percent bias is calculated as

$$\frac{(\text{observed} - \text{expected})}{\text{expected}} \times 100.$$

The nominal coverage is 95%. The R code used for the simulations is provided in the supplementary material.

6.2. Number of Imputations

In order to establish how many imputations are necessary for minimal bias and coverage near 95%, simulations were performed varying the number of imputations and assessing the differences for different percentages of missing values. Table 2 displays the breakdown of the overall percentages of missing values by each of the three types. The results are presented in Table 3 and Figure 1. Table 3 displays the estimates of the percent bias, the mean square error, and the coverage probability varying the number of imputations at each stage. Figure 1 graphically displays the change in coverage for the varying numbers of imputations for each of the four combinations of missing values.

The results of the simulation indicate that increasing the number of imputations does not significantly improve coverage, under this setup. Figure 1 shows no distinct patterns of improved coverage as either the number of imputations or the percentage of missing values changes. Table 3 shows that the percent bias is less than 1% in all cases and that the mean square error increases as the percentage of missing values increases, which is what one would expect. Therefore, increasing the number of imputations does not significantly improve the results and a small number of imputations can be used to produce efficient results, for this parameter of interest. For other parameters of interest, this may not necessarily be the case.

6.3. Order of Imputation

In order to assess whether the order of imputation impacts the bias, coverage and mean square error, a second simulation was run under the same structure as the first with the only difference being that the missing values were imputed in the following order: $MAR_2\%$, $MAR_1\%$, $MCAR\%$. That is, the missing values in the covariate were imputed last instead of first. Table 4 and Figure 2 display the results of the simulation. For simplicity, the table refers to the two different orders as $MCAR_F$ and $MCAR_L$ to mean the simulation where $MCAR\%$ is imputed first and the simulation where $MCAR\%$ is imputed last, respectively. The figure differentiates the two by using a solid line to represent the simulation where $MCAR\%$ is imputed first and a dotted line to represent the simulation where $MCAR\%$ is imputed last.

The results show that there is not a consistent change in the bias, mean square error, or coverage when the order of imputation is switched. However, there appears to be a larger range of coverage values for the case where *MCAR%* is imputed last. The figure shows no discernible pattern that prefers one order over another and the table shows that neither the bias nor the mean square error are consistently higher for one order over another. Therefore, in the situation where there is ignorability, the order of imputation does not appear to matter significantly.

7. Discussion

Three-stage multiple imputation offers flexibility which allows different types of missing values to be imputed under different models. In this paper, we introduced the methodology of multiple imputation in three stages and the associated combining rules needed for implementation. We demonstrated via simulations that we have an unbiased and efficient estimator under the assumption of ignorability. We showed that the method works well in terms of low percent bias (less than 1%), low mean square error, and good coverage. The simulations in this article explored the impact of the order and number of imputations under a specified set of conditions, most notably, ignorability. The results show that neither the order, nor the number of imputations have significant impact on the bias, mean square error, or coverage, under this set of conditions. However, this work provides a baseline framework for more complex situations and more complex assumptions imposed on the missing values.

Extensions of this work include examining the behavior of three-stage multiple imputation in a setting with nonignorable missingness, allowing for the missing values to be of a different type (for example, nonignorable uncertainty or a missing latent class), or using the flexibility of the three-stage multiple imputation design to quantify the rates of missing information attributable to each type of missing value. All of these extensions are summarized in Boyko [4], but are outside the scope of this paper.

Acknowledgement

This project was partially supported by Award Number K01MH087219 from the National Institute of Mental Health. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute of Mental Health or the National Institutes of Health.

Appendix A. Detailed derivation of Step 4

$ES^{(b)}$ follows a scaled inverse χ^2 with degrees of freedom $\nu_1 = L - 1 + \omega^{(b)}$ and scale $s_1 = (L - 1)MS_L + I^{(b)}\omega^{(b)}$ and $ES^{(w1)}$ follows a scaled inverse χ^2 with degrees of freedom $\nu_2 = L(M - 1) + \omega^{(w1)}$ and scale $s_2 = L(M - 1)MS_{M(L)} + I^{(w1)}\omega^{(w1)}\omega^{(w1)}$ and $ES^{(w2)}$ follows a scaled inverse χ^2 with degrees of freedom $\nu^3 = LM(N - 1) + \omega^{(w2)}$ and scale $s_3 = LM(N - 1)MS_{N(M)} + I^{(w2)}\omega^{(w2)}$.

Suppose x_i^{-1} , $i = 1, \dots, p$ is independently distributed as a mean square random variable with f_i degrees of freedom. Then the quantity

$$D = \frac{1 + \sum_{i=1}^p a_i}{1 + \sum_{i=1}^p a_i x_i} \quad (\text{A.1})$$

is approximately distributed as a mean square random variable with degrees of freedom

$$\nu^{-1} = \sum_{i=1}^p \left[\frac{a_i}{1 + \sum_{i=1}^p a_i} \right]^2 \frac{1}{f_i}. \quad (\text{A.2})$$

Let

$$\begin{aligned} x_1 &= \frac{ES^{(b)}(L-1+\omega^{(b)})}{(L-1)MS_L + r^{(b)}\omega^{(b)}}, \\ x_2 &= \frac{ES^{(w1)}(L(M-1)+\omega^{(w1)})}{L(M-1)MS_{M(L)} + r^{(w1)}\omega^{(w1)}}, \\ x_3 &= \frac{ES^{(w2)}(LM(N-1)+\omega^{(w2)})}{LM(N-1)MS_{N(M)} + r^{(w2)}\omega^{(w2)}}. \end{aligned} \quad (\text{A.3})$$

Under the assumption that $ES^{(b)}$, $ES^{(w1)}$, and $ES^{(w2)}$ are independent, then x_1 , x_2 , and x_3 are also independent mean square random variables.

Now, rewrite

$$\bar{U}_{LMN} + \frac{1}{MN} \left(1 + \frac{1}{L}\right) ES^{(b)} + \frac{1}{N} \left(1 - \frac{1}{N}\right) ES^{(w1)} + \left(1 - \frac{1}{N}\right) ES^{(w2)} \quad (\text{A.4})$$

as

$$\bar{U}_{LMN} \left(1 + \frac{\beta_1}{\bar{U}_{LMN}} x_1 + \frac{\beta_2}{\bar{U}_{LMN}} x_2 + \frac{\beta_3}{\bar{U}_{LMN}} x_3 \right) \quad (\text{A.5})$$

where

$$\begin{aligned} \beta_1 &= \frac{1}{MN} \left(1 + \frac{1}{L}\right) \frac{(L-1)MS_L + r^{(b)}\omega^{(b)}}{(L-1+\omega^{(b)})}, \\ \beta_2 &= \frac{1}{N} \left(1 - \frac{1}{M}\right) \frac{L(M-1)MS_{M(L)} + r^{(w1)}\omega^{(w1)}}{(L(M-1)+\omega^{(w1)})}, \\ \beta_3 &= \left(1 - \frac{1}{N}\right) \frac{LM(N-1)MS_{N(M)} + r^{(w2)}\omega^{(w2)}}{(LM(N-1)+\omega^{(w2)})}. \end{aligned} \quad (\text{A.6})$$

From the approximation as shown previously, we know that (A.5) can be approximated by a scaled inverse χ^2 distribution with scale

$$T_{LMN} = \bar{U}_{LMN} + \beta_1 + \beta_2 + \beta_3 \quad (\text{A.7})$$

and degrees of freedom

$$\nu^{-1} = \left[\frac{\beta_1}{T_{LMN}} \right]^2 (L - 1 + \omega^{(b)})^{-1} + \left[\frac{\beta_2}{T_{LMN}} \right]^2 (L(M - 1) + \omega^{(\omega 1)})^{-1} + \left[\frac{\beta_3}{T_{LMN}} \right]^2 (LM(N - 1) + \omega^{(w 2)})^{-1}.$$

Appendix B. Proof of Corollary 1

Proof. The predictive distribution (7) for Y_{mis}^A (under distinctness) is

$$\begin{aligned} P(Y_{mis}^A | Y_{obs}, M^+) &= \int \int \int \int P(\theta, \phi^+, Y_{mis}^A, Y_{mis}^B, Y_{mis}^C | Y_{obs}, M^+) d\phi^+ d\theta dY_{mis}^B dY_{mis}^C \\ &= \frac{1}{P(Y_{obs}, M^+)} \int \int \int \int P(\theta) P(Y_{com} | \theta) \times P(\phi^+) P(M^+ | Y_{com}, \phi^+) d\phi^+ d\theta dY_{mis}^B dY_{mis}^C. \end{aligned}$$

Under MAR^+ , this becomes

$$\begin{aligned} P(Y_{mis}^A | Y_{obs}, M^+) &= \frac{1}{P(Y_{obs}, M^+)} \int \int \int \int P(\theta) P(Y_{com} | \theta) \times P(\phi^+) P(M^+ | Y_{com}, \phi^+) d\phi^+ d\theta dY_{mis}^B dY_{mis}^C \\ &= \frac{P(M^+ | Y_{obs})}{P(Y_{obs}, M^+)} \int \int P(Y_{obs}, Y_{mis}^A, Y_{mis}^B, Y_{mis}^C) dY_{mis}^B dY_{mis}^C \\ &= \frac{P(M^+ | Y_{obs})}{P(Y_{obs}, M^+)} P(Y_{obs}, Y_{mis}^A) \\ &= P(Y_{mis}^A | Y_{obs}). \end{aligned}$$

Similarly, the predictive distribution (8) for Y_{mis}^B (under distinctness) is

$$\begin{aligned} P(Y_{mis}^B | Y_{obs}, Y_{mis}^A, M^+) &= \int \int \int P(\theta, \phi^+, Y_{mis}^A, Y_{mis}^B, Y_{mis}^C | Y_{obs}, M^+) d\phi^+ d\theta dY_{mis}^C \\ &= \frac{1}{P(Y_{obs}, Y_{mis}^A, M^+)} \int \int \int P(\theta) P(Y_{com} | \theta) \times P(\phi^+) P(M^+ | Y_{com}, \phi^+) d\phi^+ d\theta dY_{mis}^C. \end{aligned}$$

Under MAR^+ , this becomes

$$\begin{aligned} P(Y_{mis}^B | Y_{obs}, Y_{mis}^A, M^+) &= \frac{1}{P(Y_{obs}, Y_{mis}^A, M^+)} \int \int \int P(\theta) P(Y_{com} | \theta) \times P(\phi^+) P(M^+ | Y_{com}, \phi^+) d\phi^+ d\theta dY_{mis}^C \\ &= \frac{P(M^+ | Y_{obs}, Y_{mis}^A)}{P(Y_{obs}, Y_{mis}^A, M^+)} \int P(Y_{obs}, Y_{mis}^A, Y_{mis}^B, Y_{mis}^C) dY_{mis}^C \\ &= \frac{P(M^+ | Y_{obs}, Y_{mis}^A)}{P(Y_{obs}, Y_{mis}^A, M^+)} P(Y_{obs}, Y_{mis}^A, Y_{mis}^B) \\ &= P(Y_{mis}^B | Y_{obs}, Y_{mis}^A). \end{aligned}$$

Finally, the predictive distribution (9) for Y_{mis}^C under distinctness and MAR^+ is

$$\begin{aligned}
P(Y_{mis}^C | Y_{obs}, Y_{mis}^A, Y_{mis}^B, M^+) &= \int \int P(\theta, \phi^+, Y_{mis}^A, Y_{mis}^B, Y_{mis}^C | Y_{obs}, M^+) d\phi^+ d\theta \\
&= \frac{1}{P(Y_{obs}, Y_{mis}^A, Y_{mis}^B, M^+)} \int \int P(\theta) P(Y_{com} | \theta) \times P(\phi^+) P(M^+ | Y_{com}, \phi^+) d\phi^+ d\theta \\
&= \frac{P(M^+ | Y_{obs}, Y_{mis}^A, Y_{mis}^B)}{P(Y_{obs}, Y_{mis}^A, Y_{mis}^B, M^+)} P(Y_{obs}, Y_{mis}^A, Y_{mis}^B, Y_{mis}^C) \\
&= P(Y_{mis}^C | Y_{obs}, Y_{mis}^A, Y_{mis}^B).
\end{aligned}$$

References

1. Barnard, J. Ph. D. thesis. Department of Statistics, University of Chicago; Chicago, IL.: 1995. Cross-Match Procedures for Multiple-Imputation Inference: Bayesian Theory and Frequentist Evaluation..
2. Barnard J, Rubin DB. Small-sample degrees of freedom with multiple imputation. *Biometrika*. 1999; 86(4):948–955.
3. Belin T. Missing data: What a little can do and what researchers can do in response. *American Journal of Ophthalmology*. 2009; 148(6):820–822.
4. Boyko, J. Ph. D. thesis. Department of Statistics, University of Connecticut; Storrs, CT.: 2013. Handling Data with Three Types of Missing Values..
5. DeSouza CM, Legedza ATR, Sankoh AJ. An overview of practical approaches for handling missing data in clinical trials. *Journal of Biopharmaceutical Statistics*. 2009; 19(6):1055–1073. [PubMed: 20183464]
6. Harel O. Inferences on missing information under multiple imputation and two-stage multiple imputation. *Statistical Methodology*. 2007; 4:75–89.
7. Harel, O. Strategies for Data Analysis with Two Types of Missing Values: From Theory to Application. Lambert Academic Publishing; Saarbrücken, Germany: 2009.
8. Harel O, Pellowski J, Kalichman S. Are we missing the importance of missing values in HIV prevention randomized clinical trials? Review and recommendations. *AIDS and Behavior*. 2012; 16:1382–1393. [PubMed: 22223301]
9. Harel O, Schafer JL. Partial and latent ignorability in missing-data problems. *Biometrika*. 2009; 96(1):37–50.
10. Harel O, Zhou X-H. Multiple imputation: Review of theory, implementation, and software. *Statistics in Medicine*. 2007; 26:3057–3077. [PubMed: 17256804]
11. Kinney SK, Reiter JP. Inferences for two-stage multiple imputation for non-response. *Journal of Statistical Theory and Practice*. 2009; 3:307–318.
12. Lipsitz SR, Parzen M, Zhao LP. A degrees-of-freedom approximation in multiple imputation. *Journal of Statistical Computation and Simulation*. 2002; 72:309–318.
13. Little, RJ.; Rubin, DB. *Statistical Analysis with Missing Data*. Wiley; New York: 2002.
14. Molenberghs, G.; Kenward, MG. *Missing Data in Clinical Studies*. 1st ed.. John Wiley & Sons, Ltd.; West Sussex, England: 2007.
15. Reiter JP. Small-sample degrees of freedom for multi-component significance tests with multiple imputation for missing data. *Biometrika*. 2007; 94:502–508.
16. Reiter JP, Raghunathan TE. The multiple adaptations of multiple imputation. *Journal of the American Statistical Association*. 2007; 102:1462–1471.
17. Rubin DB. Inference and missing data. *Biometrika*. 1976; 64:581–592.
18. Rubin, DB. *Multiple Imputation for Nonresponse in Surveys*. Wiley; Hoboken, New Jersey: 1987.
19. Rubin DB. Nested multiple imputation of NMES via partially incompatible MCMC. *Statistica Neerlandica*. 2003; 57:3–18.
20. Schafer J, Novo AA. Package ‘norm’: Analysis of multivariate normal datasets with missing values. R Foundation for Statistical Computing. 2013
21. Schafer, JL. *Analysis of Incomplete Multivariate Data*. Chapman & Hall; London: 1997.

22. Schafer JL. Multiple imputation: A primer. *Statistical Methods in Medical Research*. 1999; 8(3):3–15. [PubMed: 10347857]
23. Schafer JL, Graham JW. Missing data: Our view of the state of the art. *Psychological Methods*. 2002; 7(2):147–177. [PubMed: 12090408]
24. Shen, Z. Ph. D. thesis. Department of Statistics, Harvard University; Cambridge, MA.: 2000. Nested Multiple Imputation..
25. Wagstaff DA, Harel O. A closer examination of three small-sample approximations to the multiple-imputation degrees of freedom. *The Stata Journal*. 2011; 11(3):403–419.
26. White I, Carlin J. Bias and efficiency of multiple imputation compared with complete-case analysis for missing covariate values. *Statistics in Medicine*. 2010; 28:2920–2931. [PubMed: 20842622]
27. Wood AM, White IR, Thompson SG. Are missing outcome data adequately handled? A review of published randomized controlled trials in major medical journals. *Clinical Trials*. 2004; 1(4):368–376. [PubMed: 16279275]

Figure 1.

Estimates of coverage varying number of imputations at each stage for four combinations of percentages of missing values (ignorable case)

Figure 2.

Estimates of coverage varying number of imputations at each stage for each of the four combinations of percentages of missing values. The solid lines are used to represent the simulations where *MCAR%* is imputed first and the dotted lines to represent the simulations where *MCAR%* is imputed last

Table 1

Analysis of variance table

Source	Degree of Freedom	Sum of Squares	Mean Squares	Expected Mean Squares
Between nest	$L - 1$	$MN\Sigma(\bar{Q}_{l..} - \bar{Q}_{...})^2$	$MS_L (MNB)$	$\sigma_{W_2}^2 + N\sigma_{W_1}^2 + MN\sigma_B^2$
Within nest	$L(M - 1)$	$N\Sigma\Sigma(\bar{Q}_{lm.} - \bar{Q}_{l..})^2$	$MS_{M(L)} (NW_1)$	$\sigma_{W_2}^2 + N\sigma_{W_1}^2$
Within sub-nest	$LM(N - 1)$	$\Sigma\Sigma\Sigma(\hat{Q}^{(l,m,n)} - \bar{Q}_{lm.})^2$	$MS_{N(M)} (W_2)$	$\sigma_{W_2}^2$

Table 2

Breakdown of percentages of missing values by each type (ignorable case)

Total Percentage of Missing Values	<i>MCAR</i> %	<i>MAR</i> ₁ %	<i>MAR</i> ₂ %
75	25	25	25
50	15	15	20
40	15	5	20
15	5	5	5

Table 3

Estimates of percent bias, mean square error, and coverage varying the number of imputations at each stage (ignorable case)

Percent Missing	(L,M,N)	Percent Bias (%)	Mean Square Error	Coverage
75	(10,2,2)	0.010	23.207	0.941
	(10,5,2)	0.024	23.296	0.943
	(10,5,5)	0.024	23.174	0.945
	(50,2,2)	0.007	23.178	0.937
	(50,5,2)	0.019	23.122	0.938
	(50,5,5)	0.018	23.125	0.942
	(100,2,2)	0.011	23.077	0.938
	(100,5,2)	0.021	23.042	0.942
	(100,5,5)	0.021	23.079	0.938
50	(10,2,2)	0.181	19.098	0.939
	(10,5,2)	0.141	18.912	0.941
	(10,5,5)	0.137	18.844	0.941
	(50,2,2)	0.173	18.919	0.939
	(50,5,2)	0.134	18.821	0.942
	(50,5,5)	0.131	18.812	0.941
	(100,2,2)	0.172	18.896	0.941
	(100,5,2)	0.134	18.817	0.940
	(100,5,5)	0.131	18.805	0.941
40	(10,2,2)	0.493	18.338	0.943
	(10,5,2)	0.487	18.141	0.943
	(10,5,5)	0.486	18.110	0.943
	(50,2,2)	0.498	18.071	0.945
	(50,5,2)	0.491	18.043	0.941
	(50,5,5)	0.488	18.027	0.943
	(100,2,2)	0.498	18.022	0.943
	(100,5,2)	0.492	18.002	0.944
	(100,5,5)	0.489	17.994	0.946
15	(10,2,2)	-0.112	15.081	0.940
	(10,5,2)	-0.113	15.052	0.942
	(10,5,5)	-0.112	15.074	0.941
	(50,2,2)	-0.113	15.108	0.943
	(50,5,2)	-0.111	15.093	0.942
	(50,5,5)	-0.110	15.094	0.942
	(100,2,2)	-0.112	15.080	0.944
	(100,5,2)	-0.110	15.086	0.944
	(100,5,5)	-0.109	15.084	0.944

Table 4

Results of simulations varying the number of imputations and order of imputation. The two orders are denoted as $MCAR_F$ and $MCAR_L$ to mean the simulation where $MCAR\%$ is imputed first and the simulation where $MCAR\%$ is imputed last, respectively

Percent Missing	(L,M,N)	Percent Bias (%)		Mean Square Error		Coverage	
		$MCAR_F$	$MCAR_L$	$MCAR_F$	$MCAR_L$	$MCAR_F$	$MCAR_L$
75	(10,2,2)	0.010	0.009	23.207	24.815	0.941	0.946
	(10,5,2)	0.024	0.011	23.296	24.327	0.943	0.945
	(10,5,5)	0.024	0.011	23.174	24.327	0.945	0.945
	(50,2,2)	0.007	-0.002	23.178	24.210	0.937	0.942
	(50,5,2)	0.019	0.008	23.122	24.048	0.938	0.944
	(50,5,5)	0.018	0.008	23.125	24.048	0.942	0.944
	(100,2,2)	0.011	0.003	23.077	24.000	0.938	0.944
	(100,5,2)	0.021	0.011	23.042	23.930	0.942	0.943
	(100,5,5)	0.021	0.011	23.079	23.930	0.938	0.943
50	(10,2,2)	0.181	0.307	19.098	19.920	0.939	0.944
	(10,5,2)	0.141	0.304	18.912	19.827	0.941	0.948
	(10,5,5)	0.137	0.304	18.844	19.827	0.941	0.948
	(50,2,2)	0.173	0.285	18.919	19.930	0.939	0.938
	(50,5,2)	0.134	0.281	18.821	19.877	0.942	0.936
	(50,5,5)	0.131	0.281	18.812	19.877	0.941	0.936
	(100,2,2)	0.172	0.286	18.896	19.941	0.941	0.939
	(100,5,2)	0.134	0.281	18.817	19.906	0.940	0.938
	(100,5,5)	0.131	0.281	18.805	19.906	0.941	0.938
40	(10,2,2)	0.493	0.586	18.338	19.175	0.943	0.932
	(10,5,2)	0.487	0.581	18.141	19.144	0.943	0.931
	(10,5,5)	0.486	0.581	18.110	19.144	0.943	0.931
	(50,2,2)	0.498	0.580	18.071	18.792	0.945	0.934
	(50,5,2)	0.491	0.576	18.043	18.763	0.941	0.933
	(50,5,5)	0.488	0.576	18.027	18.763	0.943	0.933
	(100,2,2)	0.498	0.576	18.022	18.644	0.943	0.933
	(100,5,2)	0.492	0.573	18.002	18.623	0.944	0.933
	(100,5,5)	0.489	0.573	17.994	18.623	0.946	0.933
15	(10,2,2)	-0.112	0.051	15.081	13.189	0.940	0.961
	(10,5,2)	-0.113	0.058	15.052	13.156	0.942	0.958
	(10,5,5)	-0.112	0.058	15.074	13.156	0.941	0.958
	(50,2,2)	-0.113	0.052	15.108	13.116	0.943	0.960
	(50,5,2)	-0.111	0.056	15.093	13.118	0.942	0.959
	(50,5,5)	-0.110	0.056	15.094	13.118	0.942	0.959
	(100,2,2)	-0.112	0.051	15.080	13.124	0.944	0.960
	(100,5,5)	-0.112	0.051	15.080	13.124	0.944	0.960

Percent Missing	(L,M,N)	Percent Bias (%)		Mean Square Error		Coverage	
		$MCAR_F$	$MCAR_L$	$MCAR_F$	$MCAR_L$	$MCAR_F$	$MCAR_L$
	(100,5,2)	-0.110	0.055	15.086	13.121	0.944	0.961
	(100,5,5)	-0.109	0.055	15.084	13.121	0.944	0.961