



Published in final edited form as:

J Off Stat. 2016 ; 32(1): 147–164. doi:10.1515/jos-2016-0007.

A Note on the Effect of Data Clustering on the Multiple-Imputation Variance Estimator: A Theoretical Addendum to Lewis et al. (2014), *JOS*

Yulei He¹, Iris Shimizu¹, Susan Schappert¹, Jianmin Xu¹, Vladislav Beresovsky¹, Diba Khan¹, Roberto Valverde¹, and Nathaniel Schenker¹

¹National Center for Health Statistics, Centers for Disease Control and Prevention, Hyattsville, MD, 20782, U.S.A.

Abstract

Multiple imputation is a popular approach to handling missing data. Although it was originally motivated by survey nonresponse problems, it has been readily applied to other data settings. However, its general behavior still remains unclear when applied to survey data with complex sample designs, including clustering. Recently, Lewis et al. (2014) compared single- and multiple-imputation analyses for certain incomplete variables in the 2008 National Ambulatory Medicare Care Survey, which has a nationally representative, multistage, and clustered sampling design. Their study results suggested that the increase of the variance estimate due to multiple imputation compared with single imputation largely disappears for estimates with large design effects. We complement their empirical research by providing some theoretical reasoning. We consider data sampled from an equally weighted, single-stage cluster design and characterize the process using a balanced, one-way normal random-effects model. Assuming that the missingness is completely at random, we derive analytic expressions for the within- and between-multiple-imputation variance estimators for the mean estimator, and thus conveniently reveal the impact of design effects on these variance estimators. We propose approximations for the fraction of missing information in clustered samples, extending previous results for simple random samples. We discuss some generalizations of this research and its practical implications for data release by statistical agencies.

Keywords

Bayesian; complex survey design; data release; exploratory data analysis; fraction of missing information; missing data

1. Introduction

Data collected for scientific research often contain missing values. For example, the National Ambulatory Medical Care Survey (NAMCS) has been conducted by the U.S. Centers for Disease Control and Prevention's National Center for Health Statistics (NCHS) since 1973.

The survey aims to provide nationally representative data on office-based physician care. The ultimate sample unit is a doctor-patient encounter, drawn systematically from the terminus of a multistage, clustered sample design. However, NAMCS has considerable item nonresponse for race, one of the key demographics used in various analyses. These missing data, if inadequately accounted for, might lead to invalid inferences and misleading policy implications.

Multiple imputation (MI) (Rubin 1987) is a popular approach to handling missing data problems. In general, MI involves replacing each missing datum with several (D) sets of plausible values drawn from a specified imputation model, resulting in several completed datasets (i.e., data with missing values filled in by imputations). Each completed dataset is analyzed separately by a standard complete-data method. The resulting inferences, including point estimates, covariance matrices, and p -values, can then be combined to formally incorporate imputation uncertainty using the formulae given in Chapter 3 of Rubin (1987) and refined in Chapter 10 of Little and Rubin (2002). See also Subsection 2.3 for more specifics. The implementation of MI in several major statistical packages, including SAS (www.sas.com), R (www.r-project.org), and STATA (www.stata.com), has made this missing data strategy increasingly popular among practitioners (Harel and Zhou 2007).

MI was originally proposed as a Bayesian, model-based approach to survey nonresponse issues (Rubin 1978). However, it has been widely applied to data of various types such as surveys, clinical trials, and observational studies. Despite its popularity, limited research has been conducted to assess the general behavior of MI for survey data with complex sample designs such as stratification and clustering. Rubin (1987, chap.4) provided some general, asymptotic arguments for the appropriateness of MI for survey data. Rubin and Schenker (1986) used data from simple random samples (SRS) as illustrations. Meng (1995) raised the issue of “uncongeniality” for MI inferences, which occurs when imputation models might be incompatible with complete-data analysis procedures. See also Kim and Shao (2014, chap.4) for further discussion of this topic. Reiter et al. (2006) demonstrated that bias can arise when complex survey-design features are not accounted for in the imputation models.

Recently, Lewis et al. (2014) applied MI to the race variable (around 30% missing) in the 2008 NAMCS and estimated race proportions at national and domain levels. They compared the variance estimates from MI and single imputation (SI), and the study results suggested that the variance increase due to MI decreases as the design effects of the estimated proportions increase. That is, estimates with larger design effects are associated with smaller increases in estimated variance after MI, despite having similar rates of missingness. Similar patterns can also be identified in simulation studies conducted by Reiter et al. (2006). It is generally expected that the variance increase due to MI is small when the rate of missingness is small. That is, there is little difference for the variance estimates between MI and SI when there is little missing data. However, the additional role played by the design effect in MI variance estimation is unclear. This phenomenon was termed “surprising” in Lewis et al. (2014), yet no convincing theoretical justification was provided. Fully understanding the rationale behind this phenomenon is important, given the increasing number of applications of MI to complex survey data (e.g., Schenker et al. 2006). This issue is also related to the emerging topics of research on conducting MI for other types of data

(e.g., clinical trials) with clustered (multilevel) structure (see van Buuren 2012, sec. 3.8 and references therein).

In this article, we aim to provide some theoretical explanation as a complement to the empirical study in Lewis et al. (2014). We elucidate the effect of data clustering on MI variance estimation by deriving algebraic expressions and discuss its practical implications. The remainder of the article is organized as follows. In Section 2, we introduce a classic one-way normal random-effects model for balanced data to characterize the data sampled from a clustered design. This model-based setup is convenient for studying the properties of MI inference. In Section 3, we derive formulae for the between- and within-imputation variance components of MI analysis for the mean estimate under this model. The variance increase due to MI is shown to decrease as clustering (design) effects increase. Approximations for the fraction of missing information are proposed. Finally, in Section 4, we propose topics for future research.

2. Method

2.1. Complete-Data Model

Complex survey designs (e.g., in NAMCS) often include multistage stratification and clustering. It is often difficult to characterize such a process using explicit models. For simplicity, we consider single-stage cluster sampling with clusters of equal sizes (Cochran 1977, chap.9). That is, a simple random sample of m clusters, each containing n elements, is drawn from M clusters in the population. We further consider a model-based representation of this sample as follows:

$$\begin{aligned} y_{ij} &= \mu + \alpha_i + \epsilon_{ij}, \quad (1) \\ \alpha_i &\stackrel{i.i.d.}{\sim} N(0, \tau^2), \\ \epsilon_{ij} &\stackrel{i.i.d.}{\sim} N(0, \sigma^2), \end{aligned}$$

for $i = 1, \dots, m, j = 1, \dots, n$, where y_{ij} is the random variable, μ is the (super) population mean, the α_i s are between-cluster random effects, and the ϵ_{ij} s represent within-cluster measurement error, means “independent and identically distributed”.

Model (1) (a balanced, one-way normal random-effects model) and its variants are frequently used in the analysis of clustered surveys (Valliant et al. 2000, chap.8). Here we use Model (1) as a basis to derive the corresponding MI variance estimators and relate them to the design effects (Kish 1965) used in survey sampling. Model 1 and its generalizations, the mixed-effects models, are also used in the emerging literature on conducting MI for clustered data not limited to surveys (e.g., see Andridge 2011 for clustered randomized trials and Schafer and Yucel 2002 for longitudinal data).

Under Model (1), $Cov(y_{ij}, y_{ij'}) = \tau^2$ for $j \neq j'$ and $j, j' \in (1, \dots, n)$, and $Cov(y_{ij}, y_{i'k}) = 0$ for $i \neq i', i, i' \in (1, \dots, m)$ and $j, k \in (1, \dots, n)$. With complete data, the typical unbiased estimator with minimum variance for μ is the overall sample mean

$\hat{\mu}_{com} = \bar{y}_{\cdot, com} = \frac{\sum_{i=1}^m \sum_{j=1}^n y_{ij}}{mn}$. The variance of the estimator is

$Var(\hat{\mu}_{com} | \tau^2, \sigma^2) = \tau^2/m + \sigma^2/mn$. Its unbiased variance estimator is $\frac{\sum_{i=1}^m (\bar{y}_{i\cdot, com} - \bar{y}_{\cdot, com})^2}{m(m-1)}$,

where $\bar{y}_{i\cdot, com} = \sum_{j=1}^n y_{ij}/n$ is the sample average at the cluster level. On the other hand, if we were to wrongly ignore the within-cluster correlation and assume that

$y_{ij} \stackrel{i.i.d.}{\sim} N(\mu, \tau^2 + \sigma^2)$, then the variance for the overall mean would be $(\tau^2 + \sigma^2)/mn$ under

the misspecified model. The ratio between the two variances is $\frac{\tau^2/m + \sigma^2/mn}{(\tau^2 + \sigma^2)/mn} = 1 + (n-1)\rho$,

where $\rho = \tau^2/(\tau^2 + \sigma^2)$ is the intraclass correlation (coefficient). From the perspective of design-based inference, the factor $1 + (n-1)\rho$ is the design effect, showing how much the variance is changed by the use of cluster sampling instead of SRS. We let $deff_{com} = 1 + (n-1)\rho$, where $deff$ denotes “design effect” as in survey statistics literature (e.g., Cochran 1977, 242; Valliant et al. 2013, 5). This design effect can also be interpreted as a model-based misspecification effect (Skinner et al. 1989, chap.2).

Note that Model (1) ignores other features in typical complex survey data such as stratification, unequal cluster sizes, as well as multistage sampling. However, the simple expression for the design effect is useful for illustrating its connection with MI variance estimation. The limitations of Model (1) are discussed in Section 4.

2.2. Missing Data

Suppose that missing data occur in the original sample. For ease of notation, we assume that within cluster i , the first r_i out of the n observations are observed. That is, y_{ij} s are observed for $i = 1, \dots, m, j = 1, \dots, r_i$; $r_i \leq n$ and missing otherwise. Following Rubin and Schenker (1986), we assume that the missingness is completely at random (MCAR) (Little and Rubin 2002) for this univariate missing data problem. This simplified assumption allows us to focus on the effect of clustering alone, excluding predictive covariates from Model (1). Under MCAR, $E(r_i) = r$ for $i = 1, \dots, m$, also implying that the missingness is unrelated to the clustering factor. See Section 4 for discussion related to a more general assumption for the nonresponse mechanism such as missing at random (MAR).

For simplicity of derivation, we let $r_i = r$ for $i = 1, \dots, m$. The rate of missingness is therefore $(n-r)/n$. Under Model 1, it is easy to verify that the grand mean of the observed data

$\hat{\mu}_{obs} = \bar{y}_{\cdot, obs} = \frac{\sum_{i=1}^m \sum_{j=1}^r y_{ij}}{mr}$ is unbiased: $E(\hat{\mu}_{obs}) = \mu$. Its variance is

$Var(\bar{y}_{\cdot, obs} | \tau^2, \sigma^2) = \tau^2/m + \sigma^2/mr$, and an unbiased variance estimator is

$\frac{\sum_{i=1}^m (\bar{y}_{i\cdot, obs} - \bar{y}_{\cdot, obs})^2}{m(m-1)}$, where $\bar{y}_{i\cdot, obs} = \sum_{j=1}^r y_{ij}/r$ is the observed-sample mean at the

cluster level. Therefore the design effect based on the observed data is

$deff_{obs} = \frac{\tau^2/m + \sigma^2/mr}{(\tau^2 + \sigma^2)/mr} = 1 + (r-1)\rho$.

2.3. A Brief Review of MI

Before we present more specifics, we briefly review the MI framework from a Bayesian model-based perspective. For an incomplete dataset $Y = \{Y_{obs}, Y_{mis}\}$, where Y_{obs} and Y_{mis} denote the observed and missing components of Y , respectively, we are interested in estimating a (scalar) population quantity Q . From the perspective of model-based inference, Q can often be treated as a superpopulation parameter in a posited model (e.g., μ in Model (1)). We further assume that the missingness is at random, which means that the probability of missingness is only related to fully observed variables or is some constant, the latter case being MCAR as a special case of MAR. According to Rubin (1987, chap.3) and Little and Rubin (2002, sec. 10.2.1), the underlying theory behind MI analysis is

$$\begin{aligned} E(Q|Y_{obs}) &= E_{Y_{mis}} E(Q|Y_{obs}, Y_{mis}), \\ \text{Var}(Q|Y_{obs}) &= \text{Var}_{Y_{mis}} E(Q|Y_{obs}, Y_{mis}) + E_{Y_{mis}} \text{Var}(Q|Y_{obs}, Y_{mis}), \end{aligned} \quad (2)$$

where Y_{mis} (the missing values for which imputations are created) are drawn from their posterior predictive distributions $P(Y_{mis}|Y_{obs})$.

In the imputation stage of MI, we draw Y_{mis} independently D times to create D completed datasets. Let \hat{Q} denote the complete-data estimate for Q . In the analysis stage, the MI

estimator for Q is $\hat{Q}_{MI} = \frac{\sum_{d=1}^D \hat{Q}_{Y_{obs}, Y_{mis}^{(d)}}}{D}$ (the average of \hat{Q} evaluated using the completed datasets). Its variance is estimated by a weighted sum of the average within-imputation variance and the between-imputation variance. That is, $\text{Var}(\hat{Q}_{MI}) = W + \left(1 + \frac{1}{D}\right)B$, where

$$\begin{aligned} W &= \frac{\sum_{d=1}^D \text{Var}(\hat{Q}_{Y_{obs}, Y_{mis}^{(d)}})}{D} \quad \text{(the average within-imputation variance), and} \\ B &= \frac{\sum_{d=1}^D (\hat{Q}_{Y_{obs}, Y_{mis}^{(d)}} - \hat{Q}_{MI})^2}{D-1} \quad \text{(the between-imputation variance).} \end{aligned}$$

The coefficient of B , that is, $1 + \frac{1}{D}$, approaches 1 as $D \rightarrow \infty$. Rubin (1987) argued that as $D \rightarrow \infty$, $\hat{Q}_{MI} \rightarrow E(Q|Y_{obs})$, $W \rightarrow E_{Y_{mis}} \text{Var}(Q|Y_{obs}, Y_{mis})$, and $B \rightarrow \text{Var}_{Y_{mis}} E(Q|Y_{obs}, Y_{mis})$.

The increase of variance due to the use of MI instead of SI (Lewis et al. 2014) can be alternatively quantified using the fraction of missing information (FMI) (Rubin 1987), a key element of MI analysis output. FMI is approximately the ratio of between-imputation variance to total variance; $FMI \approx B/(B + W)$, with the approximate equality approaching exact equality as $D \rightarrow \infty$, also termed as the population fraction of missing information (Rubin 1987, 86 and 114). It typically depends to some extent on the percent of missingness. It also depends on the analysis of interest and the extent to which the imputation model is predictive of the missing values. For example, for a univariate missing data problem with no covariate in the imputation model, the FMI for the mean estimator is approximately the rate

of missingness (Rubin 1987, 114). However, if the imputation model includes other predictive covariates, the FMI will tend to be smaller than the item nonresponse rate, reflecting the gain in precision by using these covariates.

For brevity and clarity, we mainly consider the scenario with an infinite number of imputations ($D \rightarrow \infty$). We discuss relevant issues with a finite number of imputations in Section 4.

3. MI Variance Estimators under Model (1)

3.1. The Effect of Design Effects

We aim to relate the design effect to FMI in the scenario considered in Subsections 2.1 and 2.2. Let the imputed values from the d th imputation be $y_{ij}^{*(d)}$, $i = 1, \dots, m, j = r + 1, \dots, n$ and $d = 1, \dots, D$. Then for each completed dataset $\{\{y_{ij}\}, \{y_{ij}^{*(d)}\}\}$, the completed-data estimator

is $\hat{\mu}_{com}^{(d)} = \bar{y}_{\cdot \cdot, com}^{(d)} = \frac{\sum_{i=1}^m \sum_{j=1}^r y_{ij} + \sum_{i=1}^m \sum_{j=r+1}^n y_{ij}^{*(d)}}{mn}$. The MI estimator for μ is $\hat{\mu}_{MI} = \sum_{d=1}^D \hat{\mu}_{com}^{(d)} / D$. For the d th dataset, the within-imputation variance estimator is $\frac{\sum_{i=1}^m (\bar{y}_{i \cdot, com}^{(d)} - \bar{y}_{\cdot \cdot, com}^{(d)})^2}{m(m-1)}$, where $\bar{y}_{i \cdot, com}^{(d)} = \frac{\sum_{j=1}^r y_{ij} + \sum_{j=r+1}^n y_{ij}^{*(d)}}{n}$. The between-imputation variance estimator is $\frac{\sum_{d=1}^D (\hat{\mu}_{com}^{(d)} - \hat{\mu}_{MI})^2}{D-1}$.

In the Appendix, we consider two MI scenarios, one in which τ^2 and σ^2 are known and the other in which they are unknown and require estimation that is embedded in the imputation. In both cases, it is shown that as $D \rightarrow \infty$, $E(\hat{\mu}_{MI}) = \mu$, $Var(\hat{\mu}_{MI}) \rightarrow \tau^2/m + \sigma^2/mr$, $E(W) \rightarrow \tau^2/m + \sigma^2/mn$, and $E(B) \rightarrow \sigma^2/mr - \sigma^2/mn$.

Under Model (1), the MI estimator is asymptotically equivalent to the complete-case estimator $\hat{\mu}_{obs}$ (Subsection 2.2). This is expected because of the MCAR mechanism and no predictive covariate is included in the imputation model. This is also consistent with the case of SRS (Rubin and Schenker 1986). In addition, the expected within-imputation variance $E(W)$ is asymptotically identical to $Var(\hat{\mu}_{com})$ as if data were not missing (Subsection 2.1).

This makes intuitive sense because under a correctly specified model, the imputations are expected to retain the features of the unobserved data. Therefore the completed-data statistics shall preserve the mean and variance structure of the original, complete data. One might reasonably question the necessity of MI in this case. However, the explicit expressions for $E(B)$ and $E(W)$ shed some light on the effect of clustering on MI variance estimation.

Note that as $D \rightarrow \infty$, $FMI \rightarrow \frac{E(B)}{Var(\hat{\mu}_{MI})} = \frac{\sigma^2/mr - \sigma^2/mn}{\tau^2/m + \sigma^2/mr} = \frac{\frac{n-r}{nr}(1-\rho)}{\rho + \frac{1}{r}(1-\rho)}$. Plugging in $\rho =$

$(deff_{obs} - 1)/(r - 1) = (deff_{com} - 1)/(n - 1)$ and we can show that, in the limit,

$$FMI = \frac{n-r}{n} \frac{r - def_{obs}}{(r-1)def_{obs}} = \frac{n-r}{n} \frac{n - def_{com}}{n-r + (r-1)def_{com}}.$$

Let $n \rightarrow \infty$ (so that $r \rightarrow \infty$ for a fixed missingness rate). Then

$$FMI = \frac{n-r}{n} \frac{r - def_{obs}}{(r-1)def_{obs}} = \frac{n-r}{n} \left[\frac{1}{\left(1 - \frac{1}{r}\right)def_{obs}} - \frac{1}{r-1} \right] \approx \frac{n-r}{n} \frac{1 - \frac{def_{obs}}{r}}{def_{obs}},$$

and similarly,

$$FMI \approx \frac{n-r}{n} \frac{1 - \frac{def_{com}}{n}}{\frac{n-r}{n} + \frac{r}{n}def_{com}}.$$

However, practical surveys might have more complicated designs than the one-stage cluster design that we consider. Thus it might be difficult to pinpoint n and r in those contexts. To make the derived relationship widely useful, we aim to obtain expressions that only involve the rate of missing data ($P_{mis} = \frac{n-r}{n}$) and design-effect estimates, both of which are readily available for general surveys. Therefore we consider the following simplifications:

$$FMI \approx \frac{P_{mis}}{def_{obs}}, \quad (3)$$

and

$$FMI \approx \frac{P_{mis}}{(1 - P_{mis})def_{com} + P_{mis}}, \quad (4)$$

where P_{mis} quantifies the rate of missingness.

Note that Approximations (3) and (4) can be viewed as further approximations if $def_{obs} \ll r$ and $def_{com} \ll n$ (i.e., the design effects are much less than the cluster size). Otherwise we can treat them as upper bounds which are simple to calculate. We use Approximations (3) and (4) in the numerical illustrations (Subsection 3.2) and discuss their practical use.

Moreover, the approximations are derived for clustered data, including SRS as a special case. In the latter scenario, $\rho = 0 \Rightarrow def_{obs} = def_{com} = 1$, and thus $FMI \approx P_{mis}$, matching the results stated in Rubin (1987, 114). For data with a fixed missingness rate, Approximations (3) and (4) imply that FMI decreases as def_{obs} or def_{com} increases, explaining the phenomenon identified in Lewis et al. (2014).

In the example considered in this section, the variance of the infinite- D MI estimator is a sum of the between- and within-cluster variance, that is, τ^2/m and σ^2/mr . When the intraclass correlation (or design effect) increases, the between-cluster variance dominates the within-cluster variance. Correspondingly in MI, the imputations from each cluster can be viewed as draws around the corresponding cluster average (i.e., $y_{ij}^* \sim N(\bar{y}_{i \cdot, obs}, (1 + 1/r)\sigma^2)$; see the Appendix). Thus the associated uncertainty, which is reflected by the between-imputation variance B , is only of the magnitude of the within-cluster variance σ^2 , implying that the between-imputation variance contributes little to the total variance.

Although Approximations (3) and (4) are derived under the same Model 1, their uses in more general scenarios might yield different results. Practically, Approximation (3) can be calculated using the incomplete cases, while Approximation (4) can only be calculated using imputed data (because we do not have complete data), assuming that the imputation model adequately captures the complete-data structure and relationships. It is also plausible that the approximations do not always agree when both the design and missingness mechanisms of the survey data are more complicated than what we assume in Model 1.

3.2. Numerical Illustrations

Subsection 3.1 presents some theoretical derivations under a simple one-stage clustering design. As a follow-up study to Lewis et al. (2014), we assess the practical applicability of our theoretical results (i.e., Approximations (3) and (4)) by comparing them with real-study results of Lewis et al. (2014). Since the NAMCS data have a more complicated sample design and nonresponse mechanism, we expect to see both agreements and disagreements.

Lewis et al. (2014) estimated the ratio of the standard errors between MI and SI and used it as a metric to summarize the main findings. This ratio is a monotonic transformation of FMI as $SE(\hat{\mu}_{MI})/SE(\hat{\mu}_{SI}) = 1/\sqrt{1-FMI}$. Figure 2 of Lewis et al. (2014) plots the standard error ratios against the rates of missingness for a collection of race estimates from the multiply imputed NAMCS data. Their discussion notes no clear trend in the plot, and attributes that to the variability of design effects across the different estimates. Our Figure 1 plots the ratios as a function of missingness rates across different design effects ($deff_{com}$) based on Approximation (4), but with a different plotting symbol for each design effect (symbols A through F correspond to design effects 1, 2, 5, 10, 20, and 40, respectively). For a fixed design effect, the ratio increases as the rate of missingness increases. However, when the design effect is large, the rate of increase of ratios diminishes.

In addition, Figure 3 of Lewis et al. (2014) plots the ratios against the corresponding design effects of the same collection of estimates, clearly showing an inverse relationship between the ratios and design effects. Correspondingly, our Figure 2 plots the ratio as a function of design effect across various missingness rates based on Approximation (4) (symbols A through D correspond to 5%, 10%, 20%, 30% nonresponse rates, respectively). The pattern shown from the actual estimates (Figure 3 of Lewis et al. 2014) is well mimicked here in our Figure 2: as the design effect increases, the ratios decrease and approach 1 across different missingness rates.

Approximations (3) and (4) are based on the simple Model 1 under MCAR, and we only consider the effect of intraclass correlation. The design effects from real complex surveys can be affected by other factors such as unequal weighting, stratification, and multistage sample selection. They can also be affected if the missingness mechanism is more complicated than MCAR. To assess how well the simple approximations work, we predict the ratio of standard errors using Approximations (3) and (4) and compare them with the actual estimates from the NAMCS 2008 data.

The Appendix of Lewis et al. (2014) lists the estimated standard error ratios and design effects from the MI analysis, as well as the nonresponse rates for a wide variety of race estimates. We plug the design effects and nonresponse rates into Approximation (4) and plot the predictions against the actual ratios in the left panel of Figure 3, which also includes a 45-degree line. If the approximations work well, then we would expect to see points clustered around the 45-degree line. It appears that the prediction is reasonable overall and better with smaller standard-error ratios, which likely correspond to estimates with large design effects. On the other hand, Approximation (4) works less well with smaller design effects and tends to underpredict the actual ratios. We surmise that estimates with smaller design effects are likely associated with smaller intraclass correlations, and thus the effects of other factors on the design effect cannot be simply ignored, as they are in the derivation of Approximation (4).

Furthermore, we obtain the design-effect estimates from the observed cases, plug them into Approximation (3), and plot the predicted standard error ratios against the actual ratios in the right panel of Figure 3. As noted at the end of Subsection 3.1, Approximations (3) and (4) can behave differently in more complex situations than assumed in their derivations. Correspondingly, in this example Approximation (3) performs worse than Approximation (4), showing a more severe underprediction for estimates with smaller design effects.

3.3. Practical Implications

From an analyst's/imputer's perspective, Approximations (3) and (4) have simple forms and therefore can be practically useful in exploratory analyses, given the fact that nonresponse rates are readily available and design effects from estimates in complex surveys can be easily estimated from survey statistical packages such as SUDAAN (www.rti.org). For example, before carrying out the combining step in an MI analysis, the analyst might use a singly imputed dataset to obtain point estimates for the estimands of interests and adjust their variance using Approximation (4). Even before conducting MI, an imputer might use Approximation (3) to assess the increase of variance due to MI using design-effect estimates obtained from the observed data. However, we emphasize that the use of these approximations cannot replace principled analyses of missing data (e.g., carefully planned MI and analyses as in Lewis et al. 2014 and other literature).

From a statistical agency's perspective, we recommend releasing FMI estimates for variables with considerable missingness. This is in line with Wagner (2010), which proposed to use FMI as an alternative to the nonresponse rate in data publishing. Despite the common belief that multiply imputed data should be released for public use, we note that releasing only singly imputed data still exists in practice. This might be due in part to limited resources for

data production and maintenance, as well as challenges encountered in conveying the concepts of multiple imputation to practical data users (Lewis et al. 2014). Even if multiply imputed data are released, the typical number of data copies (e.g., $D = 5$ or 10) might not be suitable if the FMI is relatively high in certain scenarios (Graham et al. 2007). Therefore, one approach would be to release multiply imputed data with a manageable number of copies to minimize the burden on resources. To compensate for the fact that these numbers might be low in certain cases, the data release could be augmented with the FMI estimates, which are obtained from a much larger D to ensure their accuracy (Harel 2007). The computational burden in obtaining such FMI estimates would be expected to be minimal with current MI software packages. Data users might be able to decide if the number of imputations released are adequate for their analyses of interest given the FMI estimate, for example, by using Rubin (1987, table 4.1) and Graham et al. (2007).

4. Discussion

In this article, we use a one-stage equal clustering sampling design and its model-based characterization to derive the variance components of the MI estimator for the mean estimand. We show that the increase in variance due to MI (or the fraction of missing information) is affected in opposite directions by the frequency of missingness and design effect. Our research is a complement to the empirical investigation in Lewis et al. (2014), one of the first studies identifying such a pattern in practice. Approximations (3) and (4) might be used as simple rules of thumb to gauge the effect of design effects on MI variance estimation.

Approximations (3) and (4) are derived assuming the number of imputations $D \rightarrow \infty$. With a finite D , we conjecture that the main pattern still holds. To see that, note that

$$FMI_D = \frac{r_D + 2(\nu_D + 3)}{r_D + 1}, \quad (5)$$

where FMI_D defines the fraction of missing information with a finite D , $r_D = (1 + D^{-1})B_D/\bar{W}_D$ and ν_D denotes the degrees of freedom in MI analysis (Rubin 1987; Barnard and Rubin 1999). As $D \rightarrow \infty$, FMI_D approaches FMI which is used in our derivation (Section 3). It could be cumbersome to plug the expressions for variance components (see Appendix) into Equation (5). On the other hand, we can gauge the impact of the design effect with finite D using a well-established large-sample result (Rubin 1987, 114): $V(\bar{Q}_D) = (1 + FMI/D)V(\bar{Q}_\infty)$, which states that the efficiency of the finite- D repeated-imputation estimator relative to the fully efficient infinite- D repeated-imputation estimator is $(1 + FMI/D)^{-1/2}$ in units of standard errors. In our scenario, we show that $FMI \rightarrow 0$ as the design effect increases. This implies that $V(\bar{Q}_D) \approx (1 + 0/D)V(\bar{Q}_\infty) = V(\bar{Q}_\infty)$ accordingly. Therefore the behavior of $V(\bar{Q}_D)$ is expected to be similar to that $V(\bar{Q}_\infty)$ with an increasing design effect.

In addition, one of the key assumptions behind the MI combining rules is that the variance of the within-imputation variance estimator is (asymptotically) much less than the between-imputation variance (Rubin 1987, 89, eq. 3.3.3). That is, $Var(\hat{W}^{(d)}) \ll E(B)$, where $\hat{W}^{(d)}$ is computed from the d th completed dataset. Note that in the scenario considered in this article, as the design effect increases, $FMI \rightarrow 0$, implying that $E(B) \rightarrow 0$. Therefore we believe that using a singly imputed dataset can reliably estimate the within-imputation variance with

moderate or large sample size. Obviously using $W = \frac{\sum_{d=1}^D \hat{W}^{(d)}}{D}$ (the average from multiply imputed datasets) would produce a more precise estimate for the within-imputation variance. However, this improvement might be minimal compared to the magnitude of the between-imputation variance. More importantly, the main need of multiply imputed data is to reliably estimate the between-imputation variance.

There are several limitations to the current research. First, the derivation assumes MCAR, which can be unrealistic. As a follow-up study to Lewis et al. (2014), the focus of this article is to elucidate the effect of clustering alone on MI variance estimation. More generally, this work can be treated as an extension of Rubin and Schenker (1986), which also focused on MCAR, to clustered data. Assuming a more plausible MAR mechanism implies accounting for the effect of predictive covariates. It is usually believed that FMI would be reduced (i.e., be less than P_{mis}) if the imputation model contained predictive covariates. However, in our limited experience, an explicit formula/relationship has not been proposed and is presumably more complicated. We are currently working on this problem.

Secondly, we conduct the derivations under a rather simplified design (model). The original NAMCS sample design involves features such as stratification and multistage sampling, leading to variable analysis weights which can also affect the design effects (Valliant et al. 2013, sec. 14.4.1 and references therein). In future research, we will study the effect of the design effect on MI estimator with unequal weighting schemes and other factors involved in complex surveys. For example, we might consider a population model (Valliant et al. 2013, 364)

$$y_{hi} \sim N(\mu_h, \sigma_h^2), \quad (6)$$

$$P(h=1) = P_1, \dots, P(h=H) = P_H, \sum_{h=1}^H P_h = 1, \quad (7)$$

where h indicates the h th stratum (or poststratum), $i = 1, \dots, n_h$ indicates the sample selected from that stratum, and P_h indicates the population fraction of the h th stratum. Under such a model, the population mean is $u = \sum_{h=1}^H P_h \mu_h$. Unequal weighting occurs when the P_h s are not all equal. We also aim to further extend our work to a more general scenario including both unequal weighting and clustering, understanding how they jointly affect the multiple-imputation variance estimation.

The current research only focuses on the population mean estimand, yet many other estimands such as regression coefficients (controlling for some covariates) are also of major interest in MI analyses. Design effects for regression coefficients have recently been studied (Lohr 2014), and thus it is of interest to include regression analyses in future studies. Furthermore, we will consider extensions to noncontinuous variables, noting that in NAMCS 2008 race is a categorical variable.

Although MI was originally proposed to handle survey nonresponse problems and has been readily applied to a wide variety of data types, systematic methods studies are lacking for understanding its behavior when applied to data with complex survey designs. Together with Lewis et al. (2014), this study can be viewed as a building block for research in this important area. In addition, further studies involving real data, such as that discussed in Lewis et al. (2014), will be invaluable for suggesting theoretical research as well as calibrating it to the real world.

Acknowledgments:

The authors would like to thank Alan Dorfman, Donald Malec, and Jennifer Madans for their valuable input.

Appendix

MI when τ^2 and σ^2 are known

We first consider the MI scheme which assumes that τ^2 and σ^2 are known. Rewrite Model 1 as

$$\begin{aligned} y_{ij} &= \mu_i + \epsilon_{ij} \\ \mu_i &\stackrel{i.i.d.}{\sim} N(\mu, \tau^2) \\ \epsilon_{ij} &\stackrel{i.i.d.}{\sim} N(0, \sigma^2). \end{aligned} \quad (1)$$

Also let the cluster-level mean (from observed data) $\bar{y}_{i \cdot, obs} = \frac{\sum_{j=1}^r y_{ij}}{r}$; then

$$Var(\bar{y}_{i \cdot, obs}) = \tau^2 + \frac{\sigma^2}{r}.$$

For simplicity, we impose a diffuse prior for μ , i.e., $p(\mu) \propto 1$. The Bayesian imputation scheme consists of

Step 1: Drawing $\mu_i^* \sim p(\mu_i | y_{obs}, \tau^2, \sigma^2)$

Step 2: Drawing $\epsilon_{ij}^* \sim N(0, \sigma^2)$

Step 3: Imputing $y_{ij}^* = \mu_i^* + \epsilon_{ij}^*$ for $i = 1, \dots, m$ and $j = r + 1, \dots, n$.

First we establish the distributions of the multiply imputed data. After some algebra, μ_i^* in

Step 1 can be expressed as $\mu_i^* = \frac{\tau^2 \bar{y}_{i \cdot, obs} + \frac{\sigma^2}{r} \bar{y}_{\cdot \cdot, obs}}{\tau^2 + \frac{\sigma^2}{r}} + a^* + b_i^*$, where $a^* \sim N\left(0, \frac{1}{m} \frac{(\sigma^2/r)^2}{\tau^2 + \sigma^2/r}\right)$

and $b_i^* \sim N\left(0, \frac{\tau^2 \sigma^2/r}{\tau^2 + \sigma^2/r}\right)$. Thus, the imputed value in Step 3 can be expressed as

$$y_{ij}^* = \frac{\tau^2 \bar{y}_{i \cdot, obs} + \frac{\sigma^2}{r} \bar{y}_{\cdot \cdot, obs}}{\tau^2 + \frac{\sigma^2}{r}} + a^* + b_i^* + \epsilon_{ij}^* \quad (2)$$

for $i = 1, \dots, m$ and $j = r + 1, \dots, n$ where $a^* \perp \{b_i^*\} \perp \{\epsilon_{ij}^*\}$, and \perp indicates independence.

Note that a^* is identical across all i s whereas the b_i^* s are different across i s.

Note that when $\tau^2 \gg \frac{\sigma^2}{r}$, $\frac{\tau^2 \bar{y}_{i \cdot, obs} + \frac{\sigma^2}{r} \bar{y}_{\cdot \cdot, obs}}{\tau^2 + \frac{\sigma^2}{r}} \approx \bar{y}_{i \cdot, obs}$, $a^* \approx 0$, $b_i^* \sim N\left(0, \frac{\sigma^2}{r}\right)$, and the

distribution of the imputed value can be approximated as $y_{ij}^* \sim N\left[\bar{y}_{i \cdot, obs}, \left(1 + \frac{1}{r}\right)\sigma^2\right]$.

By repeating the above process independently D times, we create D completed datasets in which the missing data are imputed as

$$y_{ij}^{*(d)} = \frac{\tau^2 \bar{y}_{i \cdot, obs} + \frac{\sigma^2}{r} \bar{y}_{\cdot \cdot, obs}}{\tau^2 + \frac{\sigma^2}{r}} + a^{*(d)} + b_i^{*(d)} + \epsilon_{ij}^{*(d)}, \quad (3)$$

for $d = 1, \dots, D$. Note that $a^{*(d_1)} \perp a^{*(d_2)}$, $\{b_i\}^{*(d_1)} \perp \{b_i\}^{*(d_2)}$, and $\left\{\epsilon_{ij}^{*(d_1)}\right\} \perp \left\{\epsilon_{ij}^{*(d_2)}\right\}$ for any $d_1, d_2 \in (1, \dots, D)$ and $d_1 \neq d_2$.

Secondly, we derive the forms of mean and variance estimates using Equations (1) and (3).

For the d th completed dataset $\left\{\left\{y_{ij}\right\}, \left\{y_{ij}^{*(d)}\right\}\right\}$, its mean is

$$\hat{\mu}_{com}^{(d)} = \frac{\sum_{i=1}^m \sum_{j=1}^r y_{ij} + \sum_{i=1}^m \sum_{j=r+1}^n y_{ij}^{*(d)}}{mn} \quad (4)$$

$$= \bar{y}_{\cdot \cdot, obs} + \frac{n-r}{n} \left(a^{*(d)} + \bar{b}^{*(d)} + \bar{\epsilon}_{\cdot \cdot}^{*(d)} \right) \quad (5)$$

where $\bar{b}^{*(d)} = \frac{\sum_{i=1}^m b_i^{*(d)}}{m}$, and $\bar{\epsilon}^{*(d)} = \frac{\sum_{i=1}^m \sum_{j=r+1}^n \bar{\epsilon}_{ij}^{*(d)}}{m(n-r)}$.

The MI estimator is $\hat{\mu}_{MI} = \frac{\sum_{d=1}^D \hat{\mu}_{com}^{(d)}}{D}$. Because the $\hat{\mu}_{com}^{(d)}$ s are identically distributed (yet correlated with each other), we have

$$E(\hat{\mu}_{MI}) = E(\hat{\mu}_{com}^{(d)}) = E(\bar{y}_{\cdot, obs}) = \mu; \quad (6)$$

thus μ_{MI} is unbiased.

Its variance is $Var(\hat{\mu}_{MI}) = Var\left(\frac{\sum_{d=1}^D \hat{\mu}_{com}^{(d)}}{D}\right) = \frac{D-1}{D} cov(\hat{\mu}_{com}^{(1)}, \hat{\mu}_{com}^{(2)}) + \frac{1}{D} Var(\hat{\mu}_{com}^{(1)})$. Let $D \rightarrow \infty$; then $Var(\hat{\mu}_{MI}) \rightarrow cov(\hat{\mu}_{com}^{(1)}, \hat{\mu}_{com}^{(2)}) = cov[\bar{y}_{\cdot, obs} + \frac{n-r}{n} (a^{*(1)} + \bar{b}^{*(1)} + \bar{\epsilon}^{*(1)}), \bar{y}_{\cdot, obs} + \frac{n-r}{n} (a^{*(2)} + \bar{b}^{*(2)} + \bar{\epsilon}^{*(2)})]$

After some algebra, we have as $D \rightarrow \infty$

$$Var(\hat{\mu}_{MI}) \rightarrow \tau^2/m + \frac{\sigma^2}{mr} = Var(\hat{\mu}_{obs}). \quad (7)$$

We now study the between-imputation variance $B = \frac{\sum_{d=1}^D (\hat{\mu}_{com}^{(d)} - \hat{\mu}_{MI})^2}{D-1}$. Again, because the

$\hat{\mu}_{com}^{(d)}$ s are identically distributed, $E(B) = \frac{D}{D-1} [Var(\hat{\mu}_{com}^{(1)}) - Var(\hat{\mu}_{MI})]$. As $D \rightarrow \infty$,

$E(B) \rightarrow Var(\hat{\mu}_{com}^{(1)}) - Var(\hat{\mu}_{MI})$. In addition,

$Var(\hat{\mu}_{com}^{(1)}) = Var(\bar{y}_{\cdot, obs}) + \left(\frac{n-r}{n}\right)^2 [Var(a^{*(1)}) + Var(\bar{b}^{*(1)}) + Var(\bar{\epsilon}^{*(1)})]$. Plugging in

$Var(a^{*(1)}) = \frac{1}{m} \frac{(\sigma^2/r)^2}{\tau^2 + \sigma^2/r}$, $Var(\bar{b}^{*(1)}) = \frac{1}{m} \frac{\tau^2 \sigma^2/r}{\tau^2 + \sigma^2/r}$ and $Var(\bar{\epsilon}^{*(1)}) = \frac{\sigma^2}{m(n-r)}$, we obtain

$Var(\hat{\mu}_{com}^{(1)}) = \frac{\tau^2}{m} + 2\frac{\sigma^2}{mr} - \frac{\sigma^2}{mn}$. Thus as $D \rightarrow \infty$,

$$E(B) \rightarrow \frac{\sigma^2}{mr} - \frac{\sigma^2}{mn}. \quad (8)$$

For the d th completed dataset, the within-imputation variance $W^{(d)}$ is calculated as

$\frac{\sum_{i=1}^m (\bar{y}_{i \cdot, com}^{(d)} - \bar{y}_{\cdot, com}^{(d)})^2}{m(m-1)}$. The average of the within-imputation variance is

$W = \frac{\sum_{d=1}^D W^{(d)}}{D}$. Because the $W^{(d)}$ s are identically distributed across d s, $E(W) = E(W^{(d)})$.

For simplicity, we ignore the notational index d in the following derivations. Note that

$$E(W) = E\left(\frac{\sum_{i=1}^m (\bar{y}_{i\cdot, com} - \bar{y}_{\cdot\cdot, com})^2}{m(m-1)}\right) = E\left(\frac{\bar{y}_{1\cdot, com}^2 - \bar{y}_{\cdot\cdot, com}^2}{m-1}\right) \text{ given the identical distributions}$$

of $\bar{y}_{i\cdot, com}$ across i s. In addition, $E(\bar{y}_{i\cdot, com}) = E(\bar{y}_{\cdot\cdot, com}) = \mu$, and

$$\text{Var}(\bar{y}_{\cdot\cdot, com}) = \frac{1}{m} \text{Var}(\bar{y}_{1\cdot, com}) + \frac{m-1}{m} \text{cov}(\bar{y}_{1\cdot, com}, \bar{y}_{2\cdot, com}). \text{ Therefore}$$

$$E(W) = \frac{\text{Var}(\bar{y}_{1\cdot, com}) - \text{Var}(\bar{y}_{\cdot\cdot, com})}{m-1} = \frac{\text{Var}(\bar{y}_{1\cdot, com}) - \text{cov}(\bar{y}_{1\cdot, com}, \bar{y}_{2\cdot, com})}{m}. \text{ After some algebra,}$$

$$\text{we obtain that } \text{Var}(\bar{y}_{1\cdot, com}) - \text{cov}(\bar{y}_{1\cdot, com}, \bar{y}_{2\cdot, com}) = \left(\frac{r}{n} + \frac{n-r}{n} \frac{\tau^2}{\tau^2 + \sigma^2/r}\right)^2 \text{Var}(\bar{y}_{1\cdot, obs}).$$

$$+ \left(\frac{n-r}{n}\right)^2 [\text{Var}(b_1^*) + \text{Var}(\bar{\epsilon}_1^*)]$$

$$\text{Plugging in } \text{Var}(\bar{y}_{1\cdot, obs}) = \tau^2 + \sigma^2/r, \text{Var}(\bar{y}_{\cdot\cdot, obs}) = \frac{1}{m}(\tau^2 + \sigma^2/r), \text{Var}(b_1^*) = \frac{\tau^2 \sigma^2/r}{\tau^2 + \sigma^2/r}, \text{ and}$$

$$\text{Var}(\bar{\epsilon}_1^*) = \sigma^2/(n-r), \text{ we obtain}$$

$$E(W) = \frac{\tau^2}{m} + \frac{\sigma^2}{mn}. \quad (9)$$

Based on (7), (8), and (9), we have $E(W) + E(B) \rightarrow \tau^2/m + \sigma^2/mr = \text{Var}(\hat{\mu}_{MI})$ as $D \rightarrow \infty$, consistent with Rubin's variance combination formulae.

MI when τ^2 and σ^2 are unknown

More realistically, suppose MI is conducted without knowing τ^2 and σ^2 . We impose proper prior distributions for these parameters: $p(\mu) \sim N(0, \sigma_\mu^2)$, $p(\tau^2) \sim IG(A_{\tau^2}, B_{\tau^2})$, and

$p(\sigma^2) \sim IG(A_{\sigma^2}, B_{\sigma^2})$, where IG denotes the inverse-gamma distribution. These priors are often employed in hierarchical Bayesian models (Gelman et al. 2004).

The variance components and imputations are drawn from an integrated Gibbs sampling algorithm sketched as follows:

Step 1: Draw τ^{*2} from $p(\tau^2 | y_{obs}, \mu_{\hat{r}}, \sigma^2)$;

Step 2: Draw σ^{*2} from $p(\sigma^2 | y_{obs}, \mu_{\hat{r}}, \tau^{*2})$;

Step 3: Draw μ_i^* from $p(\mu_i^* | y_{obs}, \tau^{*2}, \sigma^{*2})$.

For a single imputation, we repeat Steps 1–3 until the Gibbs chain converges. We then draw $\epsilon_{ij}^* \sim N(0, \sigma^{*2})$, and impute $y_{ij}^* = \mu_i^* + \epsilon_{ij}^*$ for $i = 1, \dots, m$ and $j = r+1, \dots, n$, where μ_i^* and σ^{*2} are the draws from the last iteration of the chain. We repeat this procedure independently D times to construct D completed datasets.

The posterior distributions of τ^2 and σ^2 under a common class of priors (including ours here) are very complicated (Box and Tiao 1973, chap.6), and therefore it is difficult to obtain their moments using explicit formulae. Nevertheless, we can assess the MI variance estimators asymptotically. In a general scenario, as stated in (Gelman et al. 2004, 587), the posterior distribution of a parameter θ approaches normality with mean θ_0 and variance $[nJ(\theta_0)]^{-1}$ as the sample size $n \rightarrow \infty$ and subject to some regularity conditions, where θ_0 is the value that minimizes the Kullback-Leibler information and J is the Fisher information. Therefore $E(\theta^*) \rightarrow \theta_0$ and $E(\theta^{*2}) \rightarrow \theta_0^2$ as $n \rightarrow \infty$, where θ^* is a draw from the posterior distribution of θ . In our context, θ can be viewed as a smooth function, say f , of τ^2 and σ^2 , θ_0 as the same function evaluated at the true τ^2 and σ^2 (from the frequentist's perspective), and sample size n can be viewed as the number of clusters m . Correspondingly, we have $E[f(\tau^{*2}, \sigma^{*2})] \rightarrow f(\tau^2, \sigma^2)$ and $E[f(\tau^{*2}, \sigma^{*2})^2] \rightarrow f^2(\tau^2, \sigma^2)$ as $m \rightarrow \infty$, where τ^{*2} and σ^{*2} are draws from the posterior distributions of τ^2 and σ^2 .

After the Gibbs sampler converges, imputation y_{ij}^* can be expressed as

$$y_{ij}^* = \frac{\tau^{*2} \bar{y}_{i \cdot, obs} + \frac{\sigma^{*2}}{r} \bar{y}_{\cdot \cdot, obs}}{\tau^{*2} + \frac{\sigma^{*2}}{r}} + a^* + b_i^* + \epsilon_{ij}^* \quad (10)$$

for $i = 1, \dots, m$ and $j = r + 1, \dots, n$, and $a^* \perp \{b_i^*\} \perp \{\epsilon_{ij}^*\}$, where $a^* \sim N\left(0, \frac{1}{m} \frac{(\sigma^{*2}/r)^2}{\tau^{*2} + \sigma^{*2}/r}\right)$,

and $b_i^* \stackrel{i.i.d.}{\sim} N\left(0, \frac{\tau^{*2} \sigma^{*2}/r}{\tau^{*2} + \sigma^{*2}/r}\right)$, and $\epsilon_{ij}^* \stackrel{i.i.d.}{\sim} N(0, \sigma^{*2})$. Here τ^{*2} and σ^{*2} are draws from their posterior distributions $p(\tau^2, \sigma^2 | y_{obs})$.

Similar to the case in which τ^2 and σ^2 are known, it is not hard to show that

$$E(\hat{\mu}_{MI}) = E(\hat{\mu}_{com}^{(d)}) = \mu. \quad (11)$$

As $D \rightarrow \infty$,

$$\text{Var}(\hat{\mu}_{MI}) \rightarrow \frac{\tau^2}{m} + \frac{\sigma^2}{mr} = \text{Var}(\bar{y}_{\cdot \cdot, obs}). \quad (12)$$

Now, the between-imputation variance is computed as $B = \frac{\sum_{d=1}^D (\hat{\mu}_{com}^{(d)} - \hat{\mu}_{MI})^2}{D-1}$. Again, because the $\hat{\mu}_{com}^{(d)}$ s are identically distributed,

$$E(B) = \frac{D}{D-1} [\text{Var}(\hat{\mu}_{com}^{(1)}) - \text{Var}(\hat{\mu}_{MI})] \rightarrow \text{Var}(\hat{\mu}_{com}^{(1)}) - \text{Var}(\hat{\mu}_{MI}) \text{ as } D \rightarrow \infty. \text{ In addition,}$$

$$\begin{aligned}
\text{Var}(\hat{\mu}_{com}^{(1)}) &= \text{Var}\left[\bar{y}_{\cdot, obs} + \frac{n-r}{n}(a^{*(1)} + \bar{b}^{*(1)} + \bar{\epsilon}^{*(1)})\right] = \text{Var}(\bar{y}_{\cdot, obs}). \text{ More specifically,} \\
&+ \left(\frac{n-r}{n}\right)^2 \left[\text{Var}(a^{*(1)}) + \text{Var}(\bar{b}^{*(1)}) + \text{Var}(\bar{\epsilon}^{*(1)})\right] \\
\text{Var}(a^*|\tau^2, \sigma^2) &= \text{Var}\left[E(a^*(1)|\tau^{*2}, \sigma^{*2})\right] + E\left[\text{Var}(a^{*(1)}|\tau^{*2}, \sigma^{*2})\right] = 0 + E\left[\frac{1}{m} \frac{(\sigma^{*2}/r)^2}{\tau^{*2} + \sigma^{*2}/r}\right] \text{ as} \\
&\rightarrow \frac{1}{m} \frac{(\sigma^2/r)^2}{\tau^2 + \sigma^2/r} \\
m \rightarrow \infty. &\text{ The last convergence holds because of the aforementioned Bayesian asymptotic} \\
&\text{arguments. Similarly, we have } \text{Var}(\bar{b}^{*(1)}|\tau^2, \sigma^2) = E\left[\frac{1}{m} \frac{\tau^{*2}\sigma^{*2}/r}{\tau^{*2} + \sigma^{*2}/r}\right] \rightarrow \frac{1}{m} \frac{\tau^2\sigma^2/r}{\tau^2 + \sigma^2/r}, \text{ and} \\
\text{Var}(\bar{\epsilon}^{*(1)}|\tau^2, \sigma^2) &= E\left[\frac{\sigma^{*2}}{m(n-r)}\right] \rightarrow \frac{\sigma^2}{m(n-r)}. \text{ This leads to}
\end{aligned}$$

$$E(B) \rightarrow \frac{\sigma^2}{mr} - \frac{\sigma^2}{mn}. \quad (13)$$

For ease of notation, we drop the conditioning on τ^2 and σ^2 in the following expressions in evaluating the within-imputation variance W and assume $m \rightarrow \infty$. Similar to the case with τ^2 and σ^2 known, we have $E(W) = \frac{\text{Var}(\bar{y}_{1 \cdot, com}) - \text{Cov}(\bar{y}_{1 \cdot, com}, \bar{y}_{2 \cdot, com})}{m}$. First,

$$\begin{aligned}
&\text{Var}(\bar{y}_{1 \cdot, com}) \text{Var}\left[\frac{r}{n}\bar{y}_{1 \cdot, obs} + \frac{n-r}{n} \frac{\tau^2}{\tau^2 + \sigma^2/r} \bar{y}_{1 \cdot, obs} + \frac{n-r}{n} \frac{\sigma_r^2 \bar{y}_{\cdot, obs}}{\tau^2 + \sigma^2/r} + \frac{n-r}{n}(a^* + b_1^* + \bar{\epsilon}_1^*)\right] \\
&= \text{Var}\left[\left(\frac{r}{n} + \frac{n-r}{n} \frac{\tau^2}{\tau^2 + \sigma^2/r}\right) \bar{y}_{1 \cdot, obs}\right] + \text{Var}\left[\left(\frac{n-r}{n} \frac{\sigma^2/r}{\tau^2 + \sigma^2/r}\right) \bar{y}_{\cdot, obs}\right] + 2^* \\
&\quad \text{Cov}\left[\left(\frac{r}{n} + \frac{n-r}{n} \frac{\tau^2}{\tau^2 + \sigma^2/r}\right) \bar{y}_{1 \cdot, obs}, \left(\frac{n-r}{n} \frac{\sigma^2/r}{\tau^2 + \sigma^2/r}\right) \bar{y}_{\cdot, obs}\right] \\
&\quad + \left(\frac{n-r}{n}\right)^2 [\text{Var}(a^*) + \text{Var}(b_1^*) + \text{Var}(\bar{\epsilon}_{1t}^*)].
\end{aligned}$$

Furthermore,

$$\begin{aligned}
 & \text{Cov} \left[\left(\frac{r}{n} + \frac{n-r}{n} \frac{\tau^{*2}}{\tau^{*2} + \sigma^{*2}/r} \right) \bar{y}_{1 \cdot, obs}, \left(\frac{r}{n} + \frac{n-r}{n} \frac{\tau^{*2}}{\tau^{*2} + \sigma^{*2}/r} \right) \bar{y}_{2 \cdot, obs} \right] \\
 &= \text{Cov} \left\{ E \left[\left(\frac{r}{n} + \frac{n-r}{n} \frac{\tau^{*2}}{\tau^{*2} + \sigma^{*2}/r} \right) \bar{y}_{1 \cdot, obs} \middle| \tau^{*2}, \sigma^{*2} \right], \right. \\
 & \quad E \left[\left(\frac{r}{n} + \frac{n-r}{n} \frac{\tau^{*2}}{\tau^{*2} + \sigma^{*2}/r} \right) \bar{y}_{2 \cdot, obs} \middle| \tau^{*2}, \sigma^{*2} \right] \left. \right\} \\
 &+ E \left\{ \text{Cov} \left[\left(\frac{r}{n} + \frac{n-r}{n} \frac{\tau^{*2}}{\tau^{*2} + \sigma^{*2}/r} \right) \bar{y}_{1 \cdot, obs}, \left(\frac{r}{n} + \frac{n-r}{n} \frac{\tau^{*2}}{\tau^{*2} + \sigma^{*2}/r} \right) \bar{y}_{2 \cdot, obs} \right] \right\} \\
 &= \text{Cov} \left[\left(\frac{r}{n} + \frac{n-r}{n} \frac{\tau^{*2}}{\tau^{*2} + \sigma^{*2}/r} \right) \mu, \left(\frac{r}{n} + \frac{n-r}{n} \frac{\tau^{*2}}{\tau^{*2} + \sigma^{*2}/r} \right) \mu \right] \\
 &= \mu^2 \text{Var} \left(\frac{r}{n} + \frac{n-r}{n} \frac{\tau^{*2}}{\tau^{*2} + \sigma^{*2}/r} \right).
 \end{aligned}$$

Finally, $\text{Var}(b_1^*) \rightarrow \frac{\tau^2 \sigma^2 / r}{\tau^2 + \sigma^2 / r}$ and $\text{Var}(\bar{e}_1^*) \rightarrow \frac{\sigma^2}{n-r}$.

Plugging all these expressions into $E(W)$, we obtain

$$E(W) \rightarrow \tau^2 / m + \sigma^2 / mn \quad (14)$$

Based on (12), (13), and (14), Rubin's variance formulae are valid asymptotically.

5. References

- Andridge RR 2011 "Quantifying the Impact of Fixed Effects Modeling of Clusters in Multiple Imputation for Cluster Randomized Trials." *Biometrical Journal* 53: 57–74. Doi: 10.1002/bimj.201000140. [PubMed: 21259309]
- Barnard J and Rubin DB. 1999 "Small-Sample Degrees of Freedom With Multiple Imputation." *Biometrika* 86: 949–955.
- Box GEP and Tiao GC. 1973 *Bayesian Inference in Statistical Analysis*. New York: Wiley.
- Cochran WG 1977 *Sampling Techniques*. New York: Wiley.
- Graham J, Olchowski A, and Gilreath T. 2007 "How Many Imputations are Really Needed? Some Practical Clarifications of Multiple Imputation Theory." *Prevention Science* 8: 206–213. Doi: 10.1007/s11121-007-0070-9. [PubMed: 17549635]
- Harel O 2007 "Inferences on Missing Information Under Multiple Imputation and Two-Stage Multiple Imputation." *Statistical Methodology* 4: 75–89. Doi: 10.1016/j.stamet.2006.03.002.
- Harel O and Zhou XH. 2007 "Multiple Imputation: Review of Theory, Implementation, and Software." *Statistics in Medicine* 26: 3057–3077. Doi: 10.1002/sim.2787. [PubMed: 17256804]
- Gelman A, Carlin JB, Stern HS, and Rubin DB. 2004 *Bayesian Data Analysis*, 2nd ed. New York: CRC Press.
- Kim JK and Shao J. 2014 *Statistical Methods for Handling Incomplete Data*. Boca Raton: CRC Press.
- Kish L 1965 *Survey Sampling*. New York: Wiley.
- Lewis T, Goldberg E, Schenker N, Beresovsky V, Schappert S, Decker S, Sonnenfeld N, and Shimizu I. 2014 "The Relative Impacts of Design Effects and Multiple Imputation on Variance Estimates: A Case Study With the 2008 National Ambulatory Medical Care Survey." *Journal of Official Statistics* 30: 147–161. Doi: 10.2478/jos-2014-0008.
- Little RJA and Rubin DB. 2002 *Statistical Analysis of Missing Data*. New York: Wiley.

- Lohr SL 2014 “Design Effects for a Regression Slope in a Cluster Sample.” *Journal of Survey Statistics and Methodology* 2: 97–125. Doi: 10.1093/jssam/smu003.
- Meng XL 1995 “Multiple Imputation With Uncongenial Sources of Input (with discussion).” *Statistical Science* 10: 538–573.
- Reiter JP, Raghunathan TE, and Kinney SK. 2006 “The Importance of Modeling the Sampling Design in Multiple Imputation for Missing Data.” *Survey Methodology* 32: 143–149.
- Rubin DB. Multiple Imputations in Sample Surveys – a Phenomenological Bayesian Approach to Nonresponse; Proceedings of the Survey Research Methods Section of the American Statistical Association; 1978. 20–34. date and place
- Rubin DB 1987 *Multiple Imputation for Nonresponse in Surveys*. New York: Wiley.
- Rubin DB and Schenker N. 1986 “Multiple Imputation for Interval Estimation From Simple Random Samples With Ignorable Nonresponse.” *Journal of the American Statistical Association* 81: 366–374. Doi: 10.1080/01621459.1986.10478280.
- Schafer JL and Yucel RM. 2002 “Computational Strategies for Multivariate Linear Mixed-Effects Models With Missing Values.” *Journal of Computational and Graphical Statistics* 11: 421–442. Doi: 10.1198/106186002760180608.
- Schenker N, Raghunathan TE, Chiu PL, Makuc DM, Zhang G, and Cohen AJ. 2006 “Multiple Imputation of Missing Income Data in the National Health Interview Survey.” *Journal of the American Statistical Association* 101: 924–933. Doi: 10.1198/016214505000001375.
- Skinner CJ, Hold D, and Smith TFM. 1989 *Analysis of Complex Surveys*. West Sussex: Wiley.
- Valliant R, Dever JA, and Kreuter F. 2013 *Practical Tools for Designing and Weighting Survey Samples*. New York: Springer.
- Valliant R, Dorfman AH, and Royall RM. 2000 *Finite Population Sampling and Inference: A Prediction Approach*. New York: Wiley.
- Van Buuren S 2012 *Flexible Imputation for Missing Data*. Boca Raton: CRC Press.
- Wagner J 2010 “The Fraction of Missing Information as a Tool for Monitoring the Quality of Survey Data”, *Public Opinion Quarterly* 74, 233–243. DOI: 10.1093/poq/nfq007.

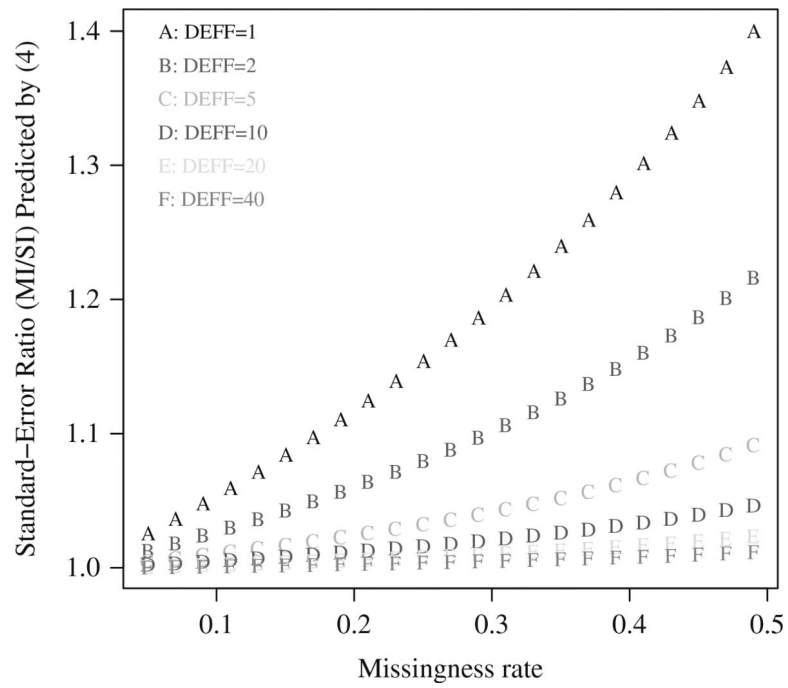


Fig. 1.

The Relationship between the Missingness Rate and Standard-Error Ratio (between MI and SI) across Different Completed-Data Design Effects. The symbols A through F correspond to design effects 1, 2, 5, 10, 20, 40, respectively

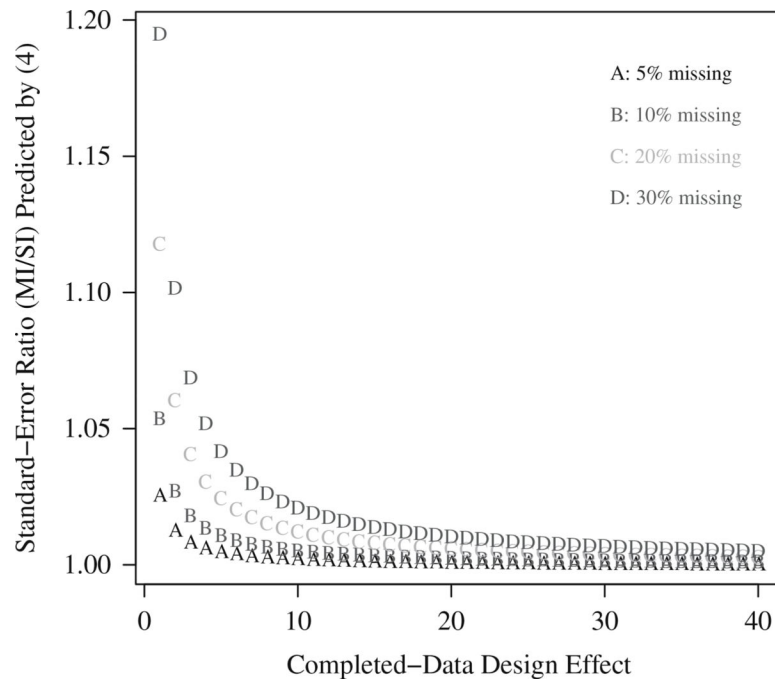


Fig. 2. The Relationship between the Completed-Data Design Effect and Standard-Error Ratio (between MI and SI) across Different Missingness Rates. The symbols A through E correspond to P_{mis} values of 5%, 10%, 20%, and 30%, respectively

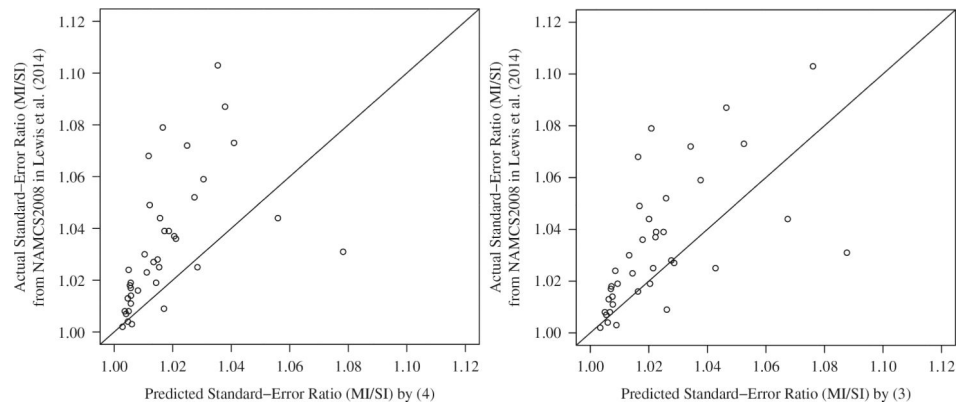


Fig. 3. The Comparison between Actual and Predicted Standard-Error Ratios. Left Panel: by Approximation (4). Right Panel: by Approximation (3)