

Multiple imputation to evaluate the impact of an assay change in national surveys

Maya Sternberg^{*†} 

National health surveys, such as the National Health and Nutrition Examination Survey, are used to monitor trends of nutritional biomarkers. These surveys try to maintain the same biomarker assay over time, but there are a variety of reasons why the assay may change. In these cases, it is important to evaluate the potential impact of a change so that any observed fluctuations in concentrations over time are not confounded by changes in the assay. To this end, a subset of stored specimens previously analyzed with the old assay is retested using the new assay. These paired data are used to estimate an adjustment equation, which is then used to 'adjust' all the old assay results and convert them into 'equivalent' units of the new assay. In this paper, we present a new way of approaching this problem using modern statistical methods designed for missing data. Using simulations, we compare the proposed multiple imputation approach with the adjustment equation approach currently in use. We also compare these approaches using real National Health and Nutrition Examination Survey data for 25-hydroxyvitamin D. Published 2017. This article is a U.S. Government work and is in the public domain in the USA.

Keywords: multiple imputation; NHANES; assay change; vitamin D

1. Introduction

Missing data is a common problem in medical studies and surveys and can arise for a myriad of reasons. A primary problem caused by missing data is the severe bias that can occur when respondents differ from non-respondents. However, not all missing data is unplanned. Graham *et al.* [1] have proposed a type of planned missingness in the setting of questionnaire design to obtain more high-quality information, without dramatically increasing the costs or the number of questions asked of each respondent. This is achieved by dividing the questions up in a specific way, ensuring a core set of questions are answered by all respondents, and using modern missing data methods to impute answers to questions that were not asked of specific sets of respondents. In our setting, when an assay has changed in a large national survey, it is cost prohibitive and often not possible because of lack of specimen availability, to go back and re-test all previously tested specimens using the new assay. This paper proposes measuring the new assay on a designated sub-sample of the originally sampled National Health and Nutrition Examination Survey (NHANES) specimens, followed by using multiple imputation to address any bias induced by changing the biomarker assay. Multiple imputation was proposed previously as a solution to bridge changes in measurement systems in large public datasets [2–5]. However, it has never been used in NHANES as a form of planned missingness to account for a change in biomarker assays.

The NHANES are a series of cross-sectional national surveys of health and nutritional status conducted on a periodic basis by the National Center for Health Statistics (NCHS) of the US Centers for Disease Control and Prevention (CDC) [6]. NHANES is the only US national survey that collects biological specimens and as such is uniquely suited to assess trends on a wide variety of health and nutritional biomarkers measured using clinical laboratory assays [7]. Ideally, for any given nutritional biomarker, the same assay is used over time so that any changes in the concentration can be evaluated

National Center for Environmental Health, Centers for Disease Control and Prevention (CDC), Atlanta, Georgia, U.S.A.

^{*}Correspondence to: Maya Sternberg, National Center for Environmental Health, Centers for Disease Control and Prevention (CDC), Atlanta, GA.

[†]E-mail: mrs7@cdc.gov

and interpreted as actual changes in the nutritional status of the US population. However, because of discontinuation or advances in laboratory methods, changes in an assay are unavoidable.

In a laboratory, it is routine to perform a method comparison study to evaluate the potential impact of a change in the assay. In the context of a national survey, if two assays are not interchangeable, the method comparison study is used to derive adjustment equations for the old assay into equivalent units of the new assay. These types of equations are published in the NHANES documentation with a recommendation that users interested in trends derive predicted values of the new assay given the old assay using the adjustment equation [8–10]. The nomenclature varies, but the adjustment equation derived from this type of method comparison study may be referred to as a crossover, bridging, or a calibration study [8–16]. In NHANES, the results of bridging studies are reported in the documentation for serum and whole blood folate, plasma homocysteine, serum ferritin, glucose, and serum 25-hydroxyvitamin D and used for analyzing trends in NHANES [8–13,17–22].

The adjustment equation derived from the bridging study is often used in NHANES to replace all the values from old assay with equivalent unit data of the new assay, and then the statistical analysis proceeds as if these were real data from the new assay. This ad hoc approach to deal with an assay change fails to account for important sources of variability, which in turn leads to underestimated standard errors, confidence intervals that are too narrow and incorrect *p*-values. Further, even though the new assay is used on a subset of the NHANES specimens to perform the bridging study, these actual measured values are not used but rather replaced themselves with predicted values. This historical approach is common practice in NHANES and was proposed as a core approach to standardize of vitamin D [15,16] and folate [23] measurements in national surveys. This paper will refer to the ad hoc approach of using the data from a bridging study to develop an equation as the adjustment equation approach.

Reframing this problem as a missing data problem, allows the use of all the statistical theory and methodology under the rubric of missing data to solve some of the shortcomings associated with the historical approach. The multiple-imputation approach has several advantages over the adjustment-equation approach. First, it provides a natural way to accommodate the uncertainty associated with replacing missing data with predictions (also called imputations). Second, this approach will not discard the actual measurements made by the new assay. Lastly, this approach places the focus on developing a proper imputation model rather than finding the single best adjustment equation from a subset of the specimens. Therefore, model fit issues such as transformations, heteroskedasticity, errors in variables, and linearity falls under the broader scope of developing a formal imputation model. Section 2 provides a background on the assays that measure serum 25-hydroxyvitamin D in NHANES and a brief review of missing data terminology and methods. Section 3 describes the proposed multiple imputation method, and for comparison, the single imputation methods considered. Section 4 illustrates the application of these imputation methods to serum 25-hydroxyvitamin D (25OHD) from NHANES 2001–2006. Section 5 outlines the details of the simulation to demonstrate how the adjustment equation approach falls short of a more principled approach. Section 6 presents some concluding remarks and a brief discussion.

2. Background

2.1. Vitamin D

Serum 25OHD is a biomarker in NHANES currently used to assess vitamin D status in the US population. A radioimmunoassay (RIA) was used to measure 25OHD in NHANES from 1988 to 1994 and 2001–2006. During these periods, the assay suffered from some bias, fluctuations, and imprecision. The CDC lab regularly uses bench and blind quality control (QC) pools that are run concurrently with patient samples to monitor the consistency of an assay's performance. Using QC pools run during 1988–2006, it was discovered that during some periods the QC pools measured higher (or lower) than in other periods. A round table was convened to address these problems [16]. The round table recommended changing the assay to an accuracy-based Liquid chromatography–tandem mass spectrometry (LC–MS/MS) method starting in 2007. In preparation for the change to the LC–MS/MS approach, a bridging study was undertaken to develop equations to adjust 25OHD measurements made with RIA to LC–MS/MS equivalent units. The details of these adjustment equations and the subsequent complete trend analysis are already published [19,20]. The adjusted data were released by NCHS on the NHANES website in October 2015 to facilitate assessing the vitamin D trends in the US population going forward [8]. In this paper, we will consider the data from the

25OHD bridging studies performed for NHANES 2001–2006 ($n = 943$). Unweighted descriptive statistics comparing RIA and LC–MS/MS for each of the survey cycles is shown in Table I, and the Bland–Altman plots are shown in Figure 1. The mean bias is statistically significant for cycles 2001–2002 (p -value < 0.0001) and 2005–2006 (p -value < 0.0001), and more importantly, the

Table I. Unweighted descriptive statistics of serum concentrations of 25-hydroxyvitamin D by radioimmunoassay and LC–MS/MS using a bridging data set: National Health and Nutrition Examination Survey 2001–2006 [20]

Survey period	<i>n</i>	Variable	Mean	SD
2001–2002	371	RIA	57.4	30.8
		LC–MS/MS	61.1	30.7
		Difference	–3.7	9.2
2003–2004	289	RIA	60.4	28.7
		LC–MS/MS	61.1	29.7
		Difference	–0.7	9.3
2005–2006	283	RIA	53.9	27.8
		LC–MS/MS	60.6	29.0
		Difference	–6.8	10.6
2001–2006	943	RIA	57.3	29.4
		LC–MS/MS	61.0	29.9
		Difference	–3.7	9.9

LC–MS/MS, Liquid chromatography–tandem mass spectrometry; RIA, radioimmunoassay.

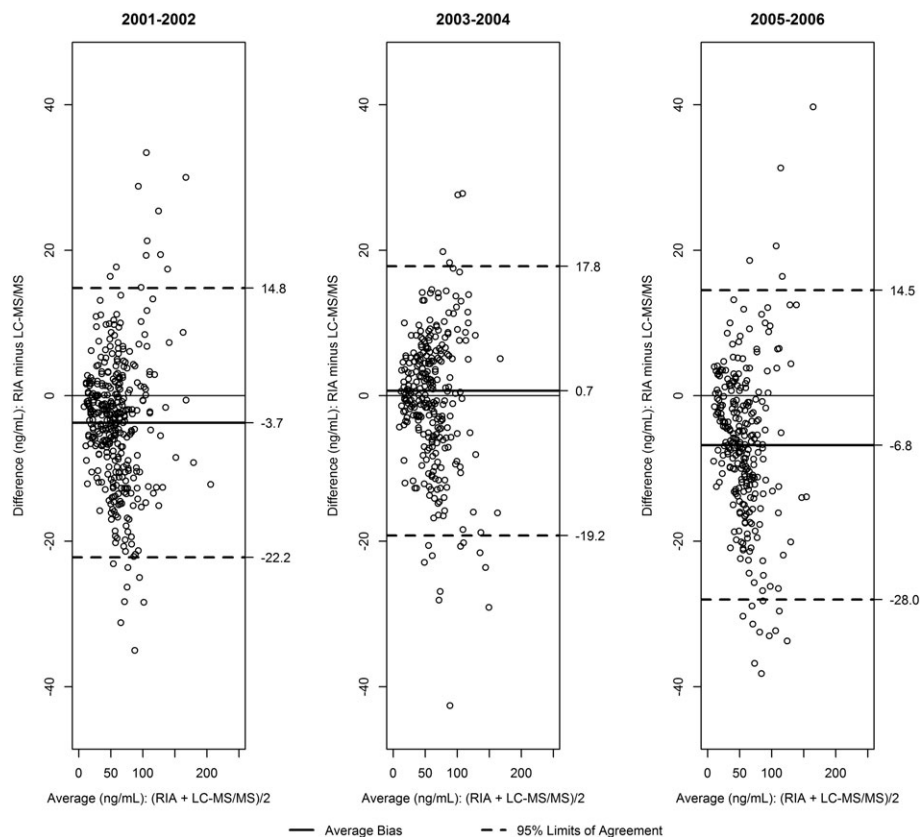


Figure 1. Bland–Altman plots comparing the serum concentrations of 25-hydroxyvitamin D by radioimmunoassay and liquid chromatography–tandem mass spectrometry (LC–MS/MS) using a bridging data set for National Health and Nutrition Examination Survey 2001–2006. RIA, radioimmunoassay.

95% limits of agreement are too wide to consider LC–MS/MS interchangeable with RIA. Because trends of the mean serum 25-OHD concentrations, reference intervals (2.5th–97.5th percentiles), and/or prevalence of vitamin D deficiency is used to assess the vitamin D status in the US population, a statistically valid approach is needed to estimate a variety of parameters under a change of assay.

2.2. Review of missing data methods

Missing data methods are commonly used to fill in the missing data with plausible value(s). This practice is often referred to as imputation. The goal for any viable imputation approach is not to predict the missing values per se but to be able to draw accurate inferences about population parameters [24]. There are many excellent reviews [24–31] and textbooks [32,33] of statistical methods to address missing data problems. Every review always starts with a discussion about the mechanism of missingness and whether the missing data is *ignorable* or *non-ignorable*. There are two ignorable missing data mechanisms: missing completely at random (MCAR) and missing at random (MAR). Rubin [32] formalized the notions of MCAR and MAR by specifying the missing data mechanism as part of the full posterior distribution in a Bayesian framework and showed that when the missing data are either MCAR or MAR, the missing data mechanism can be ignored and likelihood based approaches provide valid unbiased estimates. For this reason modern missing data methods are generally valid under MCAR or MAR.

A common approach to dealing with missing data is referred to as complete-case analysis. In other words, cases that are missing in the proposed analysis are simply dropped. This kind of approach will only provide valid estimates under MCAR, with some loss in power due to the reduction of the sample size. Imputation methods can be classified as either a single imputation or a multiple imputation method. A single imputation method replaces each missing datum with a single plausible value and typically proceeds as if the completed data are real. Some single-value imputation methods are ad hoc, such as mean imputation or last observation carried forward and known to provide biased inferences. Other single imputation methods are able to provide unbiased estimates of some population parameters under MAR but generally underestimate the standard error by ignoring the uncertainty added by imputing the missing data [33]. The adjustment equation approach used by NHANES, while not described in any textbook, can be considered an ad hoc single imputation approach that applies conditional mean imputation to impute all the values, including those actually observed. The problem with most single imputation approaches, including the adjustment equation, is that they underestimate the standard error by failing to account for two important sources of variation (i) sampling variability (ii) imputation uncertainty. Improvements on this front can be made by using a stochastic version of conditional mean imputation which adds a randomly drawn residual from the estimated error distribution of the missing data given the observed data to each predicted value. Additionally, some replication based variance estimation methods such as jackknife and bootstrapping have been proposed to address sources of uncertainty when single imputation is used [34,35].

Multiple imputation is a general term for methods that replaces a single missing value with M plausible values ($M \geq 2$). These imputed values are used to fill in the missing values and create M separate complete datasets. The M complete datasets are each analyzed separately using valid statistical procedures and the resulting estimates and standard errors are then combined to form a single inference using Rubin's formulas to approximate the posterior mean and variance of the selected parameter [32]. The variation across the M imputations reflects the uncertainty associated with replacing missing values with imputed values. The M imputations can be simulated from the posterior predictive distribution of the missing data given the observed data. This approach was used to obtain imputations under the multivariate normal distribution and applied to NHANES III data to deal with non-response, using data augmentation, a class of Markov chain Monte Carlo methods [36]. Another multiple imputation algorithm, which may have less of a theoretical basis, is to use separate chained equations, also known as sequential regression [37] or fully conditional specification [38]. In contrast to specifying the full parametric of joint distribution, it specifies each separate conditional probability distribution through a series of related (chained) regression models. Simulations suggest that this approach works well [38] and was used to deal with missing dual-energy X-ray absorptiometry data in NHANES [39].

3. Description of imputation methods for assay changes

The old assay values (X) will be completely observed by design, the new assay value (Y) will be partially observed on the basis of whether or not specimens are selected to be part of the bridging study. This type of missing data pattern is known as monotone missingness and can simplify the multiple imputation algorithm. In fact, if the joint distribution of (X, Y) is bivariate normal, a fully Bayesian approach for this problem with non-informative priors or a fully conditional specification will be asymptotically equivalent. Let $\mathbf{X} = \mathbf{X}_{\text{obs}} = (X_1, X_2, \dots, X_N)$ denote the values from the fully observed old assay on N specimens. K specimens are selected at random from N ($K < N$) and tested using the new assay, \mathbf{Y} . We denote elements \mathbf{Y} that are observed and missing as \mathbf{Y}_{obs} and \mathbf{Y}_{miss} , respectively.

3.1. Single imputation models

The estimated least squares linear regression model using the complete cases ($\mathbf{X}_{\text{obs}}, \mathbf{Y}_{\text{obs}}$) forms the basis of both single imputation models considered in this paper: the adjustment equation approach and stochastic conditional regression. Let $\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X$ be the predicted value of Y given X using the least squares estimates from the observed data. Of course, the regression form selected might include more flexible forms such as splines and/or include higher order terms, but for the purposes of this exposition we will use ordinary least squares. The adjustment equation approach uses the estimated regression equation to replace all the values of $\mathbf{Y} = (\mathbf{Y}_{\text{miss}}, \mathbf{Y}_{\text{obs}})$, thereby discarding the actual measurements of \mathbf{Y}_{obs} . In contrast, stochastic conditional mean imputation fills in the $N-K$ unknown/missing values, \mathbf{Y}_{miss} , using the estimated regression equation after adding a randomly drawn residual (r_i) from a Normal distribution with mean zero and variance equal to the estimated residual squared error of the regression line, $r_i \sim N(0, \hat{\sigma}_{y|x}^2)$; such that, the predicted value for the i th case is calculated as $\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 \hat{X}_i + r_i$.

These two single imputation approaches are compared graphically in Figure 2 using a subset of data from the bridging study to develop equations to adjust 25OHD measurements. The open circles are the observed data ($\mathbf{X}_{\text{obs}}, \mathbf{Y}_{\text{obs}}$), where the x -values are the original RIA measurements made in 2007 and the y -values are the corresponding paired LC-MS/MS measurements. The estimated least squares regression line is drawn in each graph. The solid circles along the x -axis are the selected observed values of the old assay (RIA), that need to be imputed for the new assay (LC-MS/MS). The plus signs are the imputed values made on the basis of the selected method. Notice how the adjustment equation approach replaces all the values with imputations including the original \mathbf{Y}_{obs} (Figure 2a). In stark contrast, stochastic conditional mean imputation only imputes the missing values, that is, the solid circles (Figure 2b) and includes the sampling variability around the estimated line. This graph illustrates why these single imputation approaches underestimate the variability in the new assay as they fail to capture the variability associated with the actual imputations. However, the adjustment equation approach has more drawbacks. First, as noted, the adjustment equation approach further underestimates the variance of the new assay because all the newly imputed data lie exactly along the estimated regression function. Second, discarding valid measurements made on the new assay in favor of values predicted by the regression model makes little sense and leads to a further underestimation of the variance. Third, the correlation between the two assays is severely biased to 1. Lastly, these single imputation approaches ignore the imputation uncertainty of the selected regression model. These shortcomings are resolved with a multiple imputation approach which replaces each missing value with M independent values and incorporates both the uncertainty because of the sampling variability around the line as well as the uncertainty of the parameters of the selected regression model.

3.2. Multiple imputation model

We propose a Bayesian linear regression model as the basis of the imputation model for our problem where the new assay is only partially observed. It is a common practice in laboratories to use a standard linear regression to describe the relationship between two assays such as $Y = \beta_0 + \beta_1 X + \varepsilon$, where $\varepsilon \sim N(0, \sigma_{y|x}^2)$. There may be times when additional covariates may be considered, or higher order terms, weighting, and/or transformations are needed to meet the standard assumptions of a linear regression model. These extensions can easily be incorporated into the approach described here.

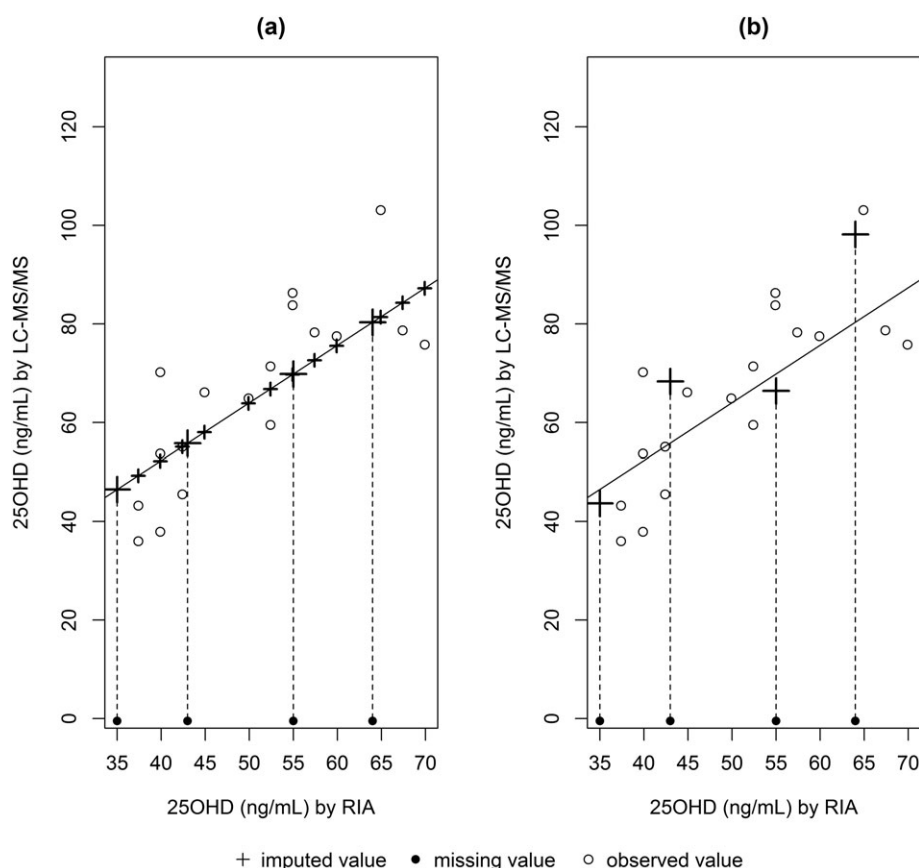


Figure 2. Single regression imputation approaches (a) adjustment equation adjustment equation (b) stochastic conditional regression. LC–MS/MS, Liquid chromatography–tandem mass spectrometry; RIA, radioimmunoassay.

Assume that (X_i, Y_i) comes from a bivariate normal distribution with mean $\mu = (\mu_x, \mu_y)$ and covariance matrix, $\Sigma = \begin{bmatrix} \sigma_x^2 & \rho\sigma_x\sigma_y \\ \rho\sigma_x\sigma_y & \sigma_y^2 \end{bmatrix}$ where σ_x^2 and σ_y^2 is the variance of x and y , respectively and ρ is the correlation between X and Y . Then the conditional distribution of Y_i given X_i is Normal with mean $\mu_{y|x} = \beta_0 + \beta_1 x$, where $\beta_0 = \mu_y - \beta_1 \mu_x$ and $\beta_1 = \rho \frac{\sigma_y}{\sigma_x}$, and variance $\sigma_{y|x}^2 = \sigma_y^2(1 - \rho^2)$.

Under a Bayesian formulation of the aforementioned simple linear regression model, we assume all the parameters, $\mu_x, \sigma_x^2, \beta_0, \beta_1, \sigma_{y|x}^2$, are random, and specify non-informative prior distributions. Start with the available cases to get initial values $\hat{\beta}_0^{(0)}, \hat{\beta}_1^{(0)}, \hat{\sigma}_{y|x}^{2(0)}$ via least squares, followed by an iterative algorithm, known as data augmentation, which cycles between an imputation step (I-step) and posterior step (P-step). At the j th I-step, we simulate the missing values from the conditional predictive distribution $Y_{miss,i}^{(j)} | x_i, Y_{obs}, \beta_0^{(j-1)}, \beta_1^{(j-1)}, \sigma_{y|x}^{2(j-1)}$. The P-step involves drawing from the conditional posterior distributions of $\hat{\sigma}_{y|x}^{2(j)}, \hat{\beta}_1^{(j)}, \hat{\beta}_0^{(j)} | Y_{miss}^{(j)}, Y_{obs}$ to obtain updated estimates $\hat{\beta}_0^{(j)}, \hat{\beta}_1^{(j)}, \hat{\sigma}_{y|x}^{2(j)}$. Under the Normal Bayesian linear model, the conditional posterior distributions of $\hat{\sigma}_{y|x}^{2(j)}, \hat{\beta}_1^{(j)}, \hat{\beta}_0^{(j)} | Y_{miss}^{(j)}, Y_{obs}$ have well-known probability distributions. Using $\hat{\beta}_0^{(j)}, \hat{\beta}_1^{(j)}, \hat{\sigma}_{y|x}^{2(j)}$, we obtain the i th subject's missing value by drawing from the conditional predictive distribution $Y_{miss,i} | x_i, Y_{obs}, \beta_0, \beta_1, \sigma_{y|x}^2$ (see Appendix for details). In order to get M imputations for the i th subject with the new assay, these steps are repeated for $j = 1 \dots M$.

Rubin [32] provided the theoretical foundation for multiple imputation using a Bayesian formulation. Using these M imputations, M separate 'filled in' datasets are created and each of these datasets is analyzed separately using a statistical analysis as if there was no missing data. The last stage combines these M separate results using Rubin's formulas into a single inference. Using the

conventional notation, let Q be the parameter of interest, such as a mean, prevalence, or percentile, so that \hat{Q}_i is the point estimate and V_i the estimated variance of \hat{Q}_i from the i th ‘filled in’ data set for $i = 1 \dots M$. Rubin [32] shows that the complete data posterior mean of Q can be found by averaging the individual estimates, $\bar{Q} = \frac{1}{M} \sum_{i=1}^M \hat{Q}_i$. Using the \hat{Q}_i ’s, we can compute the between imputation variance $B = \frac{1}{(M-1)} \sum_{i=1}^M (\hat{Q}_i - \bar{Q})^2$ and the average of the estimated variances $\bar{V} = \frac{1}{M} \sum_{i=1}^M V_i$ as the within imputation variance to obtain an estimate of the standard error for \bar{Q} as $\sqrt{(1 + M^{-1})B + \bar{V}}$. The relative increase in the variance due to the imputation can be calculated as $r = \frac{(1 + M^{-1})B}{\bar{V}}$. The value of r can be used to construct $\lambda = \frac{r + \frac{2}{3 + (M-1)(1+r^{-1})^2}}{r+1}$, the fraction of missing data. Additionally, r is used to compute the approximate degrees of freedom. Let df be the appropriate degrees of freedom associated with the variance of the complete data for a given estimator. When the degrees of freedom is small, the following formula is recommended to calculate the degrees of freedom [40]

$$\left(\frac{1}{(M-1)(1+r^{-1})^2} + \frac{1}{\left(\frac{df+1}{df+3} \right) \left[1 - \left(\frac{r\bar{V}}{\bar{V} + (1+M^{-1})B} \right) \right] df} \right)^{-1}.$$

4. Application serum total 25-hydroxyvitamin D (25OHD) in National Health and Nutrition Examination Survey

In this section, we compare the adjustment equation approach and stochastic conditional regression with multiple imputation. The adjusted 25OHD data released on the NHANES website in October of 2015 are based on the previously published adjustment equations [8,19,20]. These equations are reproduced, along with the respective estimated mean square error (MSE) and elements of the variance–covariance matrix of the coefficients in Table II for each 2-year cycle from 2001 to 2006. (SAS Institute, Carey, NC, US) v 9.3 [41] was used to perform both stochastic conditional regression (rannor statement) and the multiple imputations. To be conservative, a large number of multiple imputations are used ($M = 40$). Multiple imputations are performed separately for each cycle, similar to the adjustment equations developed by NCHS [19,20]. Two distinct multiple imputation models are considered. MI model 1 uses the measurements made from the completely observed RIA and partially observed LCMS/MS. MI Model 2 adds covariates to MI model 1. The covariates are age, race (four dummy variables), gender, and the 2-year survey mobile examination center (MEC) weight. SAS-callable SUDAAN 11.0.0 [42] is used to estimate the weighted mean 25OHD concentration and the weighted prevalence of deficiency for each of the imputation methods described using the MEC weights. Taylor series linearization for variance estimation is used to account for the stratified cluster design.

Table II. Estimated linear regression models from a bridging study to predict serum concentrations of LC–MS/MS-equivalent 25-hydroxyvitamin D from original radioimmunoassay 25-hydroxyvitamin D for National Health and Nutrition Examination Survey, 2001–2006 [19,20].

Survey Period	n	Equation	Mean Square Error	Variance of intercept	Variance of slope	Covariance of intercept and slope
2001–2001	371	LC–MS/MSequivalent = 6.43435 + 0.95212 * RIAoriginal	83.38	1.00689	0.00024	–0.01362
2003–2004	289	LC–MS/MSequivalent = 1.72786 + 0.98284 * RIAoriginal	85.88	1.61349	0.00036	–0.02179
2005–2006	283	LC–MS/MSequivalent = 8.36753 + 0.97012 * RIAoriginal	112.38	1.89393	0.00052	–0.02779

LC-MS/MS, Liquid chromatography–tandem mass spectrometry; RIA, radioimmunoassay.

The results in Tables III and IV are limited to persons 12 years of age and older, stratified by survey cycle (2001–2002, 2003–2004, and 2005–2006) and grouped by demographic variables or vitamin D supplement use. The estimated weighted means are fairly stable regardless of the imputation method. The magnitude of the increase in the standard error under multiple imputation models relative to the adjustment equation approach varies depending on the subgroup and year, and for MI Model 1 ranges from around 1%–44% (median ~ 12%), further confirming inferences based on the adjustment equation are too liberal. Table IV presents the prevalence of LC–MS/MS-equivalent serum 25OHD concentrations below the 30 nmol/L cutoff. There is a noticeable difference between the imputation methods when it comes to estimating characteristics around the tails. The estimates of the prevalence are similar for both multiple imputation models and stochastic conditional regression but differ from those based on the adjustment equation. Additionally, there can be a severe underestimation of the standard error with the adjustment equation approach, which leads to confidence intervals that are too narrow for the stated coverage. The multiple imputation (MI model 1) standard errors are 5%–190% (median ~ 40%) larger than the adjustment equation approach. The magnitude depends on the subgroup, year, and the magnitude of the prevalence itself; such as when the prevalence is low. Lastly, except for a few subgroups, the addition of the selected covariates to MI Model 1 appears to have little to no effect on the mean 25OHD concentration and prevalence <30 nmol/L estimates as compared with MI model 1 alone. The exceptions are in the 2005–2006 survey cycle: 12–19 year olds, Non-Hispanic Black and Non-Hispanic Whites.

5. Simulation study

5.1. Details of simulation study

The primary challenge to our problem is that the bridging study is based on a rather small subset of the original data, usually between 5 and 10% of the original sample. We perform simulations to evaluate the conditions under which multiple imputation can provide reasonable statistical inference for a variety of parameters in our setting. Factors that affect the magnitude of the missing data uncertainty include the percent of missing data, the predictive ability of observed (or utilized) information to make accurate imputations, and the number of imputations M . The relative efficiency of selecting M imputations over an infinite number of imputations can be estimated as $(1 + \frac{\lambda}{M})^{-1}$, where λ is the fraction of missing data [32]. Using this formula, it can be shown if $\lambda \leq 0.50$, that $M = 10$ has a relative efficiency ≥ 0.95 . For higher rates of missing information ($\lambda > 0.50$), this formula suggests selecting a larger value of M . We selected $M = 25$ to be conservative because our focus is on cases with a high percent of missing data, although the fraction of missing information (λ) may be much smaller than the actual percent of missing data when the two assays are highly correlated. M is also related to the degrees of freedom, a larger choice of M helps minimize the complexities that can arise from limited degrees of freedom. To assess these issues, we performed simulations under the following three scenarios:

- (1) X and Y have a bivariate normal distribution with mean $\mu = (\mu_x = 50, \mu_y = 60)$, covariance matrix $\Sigma = \begin{bmatrix} 190 & 190 \\ 190 & 225 \end{bmatrix}$ and correlation is 0.9189;
- (2) X and Y have a bivariate normal distribution with mean $\mu = (\mu_x = 50, \mu_y = 60)$, covariance $\Sigma = \begin{bmatrix} 190 & 103 \\ 103 & 225 \end{bmatrix}$ and correlation is 0.4982;
- (3) \sqrt{X} and \sqrt{Y} have a bivariate normal distribution with mean $\mu = (\mu_x = 7.1, \mu_y = 7.6)$, covariance matrix $\Sigma = \begin{bmatrix} 3.24 & 3.3 \\ 3.3 & 4 \end{bmatrix}$ and correlation is 0.9167. The detransformed values of X and Y are obtained by squaring the simulated data. It can be shown that the expected value of Y and X are $\mu_y^2 + \sigma_y^2 = (7.6)^2 + 4 = 61.76$ and $\mu_x^2 + \sigma_x^2 = (7.1)^2 + 3.24 = 53.65$, respectively [43].

The population means and correlation used in scenarios 1 and 3 are motivated from the bridging study for serum 25OHD using a subset of data from NHANES 2001–2006. Prevalence of vitamin D deficiency

Table III. Weighted mean LC–MS/MS-equivalent serum 25-hydroxyvitamin D concentrations for person 12 y of age and older, stratified by survey, and grouped by demographic variables or vitamin D supplement use, NHANES 2001–2006: A comparison of imputation methods.

Group	2001–2002		2003–2004		2005–2006	
	Mean	SE	Mean	SE	Mean	SE
All						
MI Model 1	62.27	0.98	62.64	1.70	60.85	1.38
MI Model 2	62.63	1.14	63.30	1.74	61.70	1.37
Stochastic CR	62.17	0.90	62.74	1.55	61.14	1.11
AE	62.25	0.88	62.72	1.62	61.00	1.12
Age, y						
12–19						
MI Model 1	62.97	1.13	63.77	2.20	61.92	1.84
MI Model 2	63.19	1.27	64.24	2.22	64.27	1.68
Stochastic CR	62.99	1.10	64.34	2.18	62.34	1.72
AE	63.00	1.04	63.91	2.10	61.92	1.62
20–39						
MI Model 1	62.80	1.13	62.85	1.97	62.35	1.72
MI Model 2	62.90	1.39	63.28	1.99	62.91	1.68
Stochastic CR	62.50	1.06	62.94	1.81	62.68	1.41
AE	62.76	1.02	62.93	1.84	62.53	1.44
40–59						
MI Model 1	62.35	1.28	62.12	2.01	59.90	1.47
MI Model 2	62.91	1.49	62.98	2.16	60.27	1.68
Stochastic CR	62.44	1.22	62.22	1.88	60.09	1.25
AE	62.38	1.14	62.18	1.98	60.12	1.15
≥60						
MI Model 1	60.55	1.31	62.40	1.40	59.35	1.40
MI Model 2	61.12	1.56	63.21	1.64	60.34	1.65
Stochastic CR	60.41	1.17	62.25	1.19	59.68	1.06
AE	60.44	1.16	62.49	1.17	59.45	1.14
Sex						
Males						
MI Model 1	63.24	1.05	62.93	1.78	60.93	1.36
MI Model 2	64.24	1.23	62.97	1.95	62.20	1.43
Stochastic CR	63.43	0.96	62.90	1.64	61.33	1.17
AE	63.22	0.90	63.07	1.71	61.06	1.01
Females						
MI Model 1	61.36	1.13	62.36	1.69	60.77	1.52
MI Model 2	61.13	1.38	63.61	1.72	61.23	1.67
Stochastic CR	60.99	1.06	62.60	1.53	60.96	1.25
AE	61.34	1.04	62.39	1.58	60.95	1.30
Race-ethnicity						
Mexican American						
MI Model 1	55.07	1.54	54.10	1.78	50.97	2.11
MI Model 2	55.64	2.12	54.10	2.06	51.08	1.37
Stochastic CR	55.40	1.53	54.15	1.73	51.21	1.86
AE	55.06	1.44	54.23	1.62	51.17	1.87
Non-Hispanic Black						
MI Model 1	39.28	0.79	40.87	1.74	41.40	1.34
MI Model 2	37.57	1.01	40.88	1.65	39.17	1.66
Stochastic CR	38.97	0.59	41.01	1.57	41.28	1.22
AE	39.34	0.55	40.91	1.48	41.67	1.03
Non-Hispanic White						
MI Model 1	67.36	1.11	68.33	1.69	66.13	1.28
MI Model 2	68.57	1.24	69.51	1.78	68.13	1.32
Stochastic CR	67.35	1.02	68.46	1.57	66.50	0.99
AE	67.33	0.99	68.40	1.63	66.23	0.99

(Continues)

Table III. (Continued)

Group	2001–2002		2003–2004		2005–2006	
	Mean	SE	Mean	SE	Mean	SE
Vitamin D supplement use						
Yes						
MI Model 1	68.15	1.13	68.64	1.60	66.60	1.43
MI Model 2	68.44	1.28	69.42	1.69	67.15	1.47
Stochastic CR	67.92	0.97	68.42	1.38	67.02	1.06
AE	68.03	0.96	68.62	1.46	66.74	1.12
No						
MI Model 1	58.95	1.03	59.08	1.82	57.21	1.51
MI Model 2	59.57	1.18	59.77	1.84	58.33	1.51
Stochastic CR	58.94	0.98	59.33	1.71	57.42	1.26
AE	58.98	0.93	59.21	1.75	57.38	1.28

Abbreviations: SE, standard error; MI model 1, multiple imputation with assay data only; MI model 2, multiple imputation with assay data, age, race, sex, and MEC survey weight; Stochastic CR, stochastic conditional regression; AE, adjustment equation.

LC–MS/MS, Liquid chromatography–tandem mass spectrometry; NHANES, National Health and Nutrition Examination Survey.

is often a parameter of interest. To investigate the behavior of this estimator, simulated values <30 were considered deficient [44]. Scenario 3 is probably most consistent with serum 25OHD values because these measurements are often approximately normal after a square root transformation [19].

Two different missing data mechanisms are considered for the Y variable. The first is based on a Bernoulli distribution. The probability a datum is missing, π , ranged from 0.5 to 0.95. Thus, the sample size of the simulated bridging study ranged from 5% to 50% of the original sample size. The other missing data mechanism is based on a systematic random sample. The sample is selected by sorting the X values and sampling across the full range of the X with an appropriate sampling interval such that 50%–95% of the values are missing. To ensure that the full range of the X measurement is included, this latter sampling approach is consistent with how calibration curves are developed for assays in clinical laboratories and the sampling strategy for previous bridging studies in NHANES. The simulation and analysis was implemented in R [45] using the package MICE [46]. The multiple imputation in MICE was performed with Bayesian linear regression under non-informative Jeffrey prior, bivariate normality, and monotone missingness. We perform 1000 simulations of each scenario where each complete dataset has 2000 observations and $M = 25$ multiple imputations. We selected a large sample size N , as about 5000 subjects of all ages are sampled in NHANES each year [7].

5.2. Results of the simulation study

We compared the results of the imputation methods described in Section 3 for a variety of estimands commonly used to assess nutritional status using biomarkers in NHANES [22]. The estimators for the following parameters are assessed: population mean, population reference range (population percentiles 2.5th and 97.5th), and population prevalence of deficiency (<30). The 95% confidence intervals for the percentiles are calculated using the Woodruff method [47]. For each imputation method, Tables V–VII (and Tables S1 and S2) present the average across the 1000 simulations of each selected estimate, the width of the 95% confidence interval and coverage of the confidence interval for various levels of missing data. Each table presents averages based on the full original data before missing data mechanism is applied to Y, complete case analysis, adjustment equation (AE), stochastic conditional regression (stochastic CR), and multiple imputation (MI). In the case of simulation scenario 3 where the detransformed data is not normal, the results are based on two different imputation models. The first model addresses the non-normality of the data by taking a square root transform of the detransformed data to develop the imputation model and then back-transforming (squaring) the imputations. The second model simply proceeds with the detransformed data.

The purpose of presenting the results from the full data set before applying a missing data mechanism and the complete case results is to set expectations to compare the imputation results. With a sample size of 2000, we expect the average coverage of the 95% confidence intervals for the estimators of the full data

Table IV. Weighted prevalence of LC–MS/MS-equivalent serum 25-hydroxyvitamin D concentrations below or above 30 nmol/L for person 12 y of age and older, stratified by survey, and grouped by demographic variables or vitamin D supplement use, NHANES 2001–2006: a comparison of imputation methods.

Group	2001–2002		2003–2004		2005–2006	
	Mean	SE	Mean	SE	Mean	SE
Overall						
MI Model 1	7.89	0.86	9.29	1.43	8.82	1.19
MI Model 2	8.04	1.02	9.03	1.43	8.62	1.26
Stochastic CR	7.78	0.71	9.23	1.29	8.65	0.89
AE	5.37	0.67	7.48	1.18	5.16	0.71
Age, y						
12–19						
MI Model 1	6.59	1.17	8.09	1.58	8.56	1.58
MI Model 2	7.13	1.28	8.02	1.61	7.90	1.56
Stochastic CR	6.26	1.16	8.23	1.30	8.67	1.48
AE	4.11	0.85	6.46	1.30	5.26	1.12
20–39						
MI Model 1	8.28	1.10	10.18	1.78	8.71	1.40
MI Model 2	8.63	1.27	10.05	1.82	8.84	1.52
Stochastic CR	8.39	0.86	10.60	1.59	8.40	1.05
AE	5.99	0.73	8.23	1.40	5.10	0.79
40–59						
MI Model 1	8.07	1.10	9.21	1.69	9.11	1.56
MI Model 2	8.00	1.29	8.94	1.65	9.17	1.66
Stochastic CR	7.65	1.06	8.49	1.61	9.21	1.14
AE	5.74	0.93	7.07	1.45	5.52	0.90
≥60						
MI Model 1	7.78	1.33	8.72	1.31	8.64	1.47
MI Model 2	7.67	1.41	8.16	1.36	7.83	1.47
Stochastic CR	7.94	0.75	8.87	0.73	8.07	0.98
AE	4.41	0.61	7.59	1.08	4.60	0.73
Sex						
Males						
MI Model 1	5.98	0.80	7.18	1.40	7.74	1.21
MI Model 2	5.81	0.84	7.52	1.52	7.41	1.30
Stochastic CR	6.05	0.62	7.31	1.28	7.86	1.02
AE	3.76	0.39	5.14	1.10	4.06	0.61
Females						
MI Model 1	9.70	1.12	11.29	1.58	9.83	1.41
MI Model 2	10.11	1.41	10.45	1.55	9.77	1.58
Stochastic CR	9.40	0.93	11.05	1.33	9.38	0.96
AE	6.89	1.07	9.69	1.35	6.20	0.94
Race-ethnicity						
Mexican American						
MI Model 1	9.31	1.32	12.55	2.28	13.62	2.61
MI Model 2	8.28	1.99	12.22	2.66	13.13	2.94
Stochastic CR	6.85	0.62	14.10	2.41	13.72	2.66
AE	5.26	0.76	9.32	1.56	7.69	2.04
Non-Hispanic Black						
MI Model 1	32.63	2.20	31.38	3.76	28.93	2.87
MI Model 2	35.54	2.78	31.47	3.42	33.15	3.86
Stochastic CR	32.63	1.38	30.22	3.04	29.40	2.79
AE	28.46	1.90	29.57	3.41	21.79	2.19
Non-Hispanic White						
MI Model 1	3.96	0.60	4.93	0.84	4.59	0.84
MI Model 2	3.19	0.65	4.28	0.85	3.22	0.73
Stochastic CR	4.07	0.44	4.74	0.74	4.35	0.50
AE	2.17	0.29	3.50	0.56	2.11	0.29

(Continues)

Table IV. (Continued)

Group	2001–2002		2003–2004		2005–2006	
	Mean	SE	Mean	SE	Mean	SE
Supplement use						
Yes						
MI Model 1	3.06	0.63	4.22	0.94	3.77	0.87
MI Model 2	3.25	0.66	4.11	0.95	3.89	0.99
Stochastic CR	2.48	0.37	3.68	0.67	3.57	0.69
AE	1.69	0.33	2.35	0.67	1.16	0.33
No						
MI Model 1	10.61	1.20	12.25	1.78	11.97	1.54
MI Model 2	10.62	1.37	11.82	1.77	11.53	1.56
Stochastic CR	10.73	1.03	12.50	1.64	11.84	1.09
AE	7.38	0.97	10.48	1.46	7.66	0.96

Abbreviations: SE, standard error; MI model 1, multiple imputation with assay data only; MI model 2, multiple imputation with assay data, age, race, sex, and MEC survey weight; Stochastic CR, stochastic conditional regression; AE, adjustment equation.

LC–MS/MS, Liquid chromatography–tandem mass spectrometry; NHANES, National Health and Nutrition Examination Survey.

set to be close to 95% and the estimates to be unbiased. We have similar expectations for the complete case approach under MCAR, although as the percent of missing data increases, we see an increase in the average width of the confidence interval in order to maintain the 95% coverage. In addition, under the complete case approach as the percent missing data increases (i.e., observed sample size gets smaller), the asymptotic assumptions associated with the confidence intervals for some estimators breaks down, for example, prevalence not having exact 95% coverage. It is also worth noting that the complete case approach overestimates the sampling error (the traditional formula for the variance) when the sampling frame is ordered under systematic random sampling and the order is highly correlated with the outcome of interest [48]. While there is no unbiased way to estimate the variance from a systematic sample of an ordered list, it is clear that the sampling variability of the sample mean is smaller than under simple random sampling, and this is reflected by having higher than expected coverage (>95%) of the confidence intervals, especially when the order of the sampling frame is highly correlated with the new assay. Figure 3 compares the sampling distribution of the sample mean from data simulated from scenario 1 when there is no missing data or when only the complete cases are used (90% of the data is missing). As expected, the sampling error associated with having no missing data (and therefore a larger sample size) is the smallest, and the sampling error associated with the complete cases selected under the systematic random sampling is much smaller than under the simple random sampling.

Regardless of the percent of data missing, all the imputation methods provide an unbiased estimate of the population mean (Table V). On the other hand, the coverage of the adjustment equation's 95% confidence intervals is well below the expected 95% because the variability of the sample mean (i.e., the standard error) is severely underestimated. In fact, the adjustment equations' average width of the 95% confidence interval is routinely smaller than the full-data scenario. This undesirable feature demonstrates why the adjustment equation is not a viable imputation approach when any statistical inference is a primary goal of the analysis. The severity of the undercoverage for both the adjustment equation and stochastic conditional regression methods is related to the percent of data that is missing, as well as the underlying true correlation between X and Y. These findings are consistent with the literature [33].

Estimates at the tails such as the prevalence of less than 30 (Table VI), the 2.5th (Table VII), and 97.5th (Table S1) percentiles are always biased using the adjustment equation. Even the 25th and 75th percentiles are biased (results not shown). However, it appears that this bias may be remedied with the stochastic conditional regression approach even when the correlation is weaker. The coverage of the 95% confidence intervals continues to be poor for both these single imputation approaches because they underestimate the uncertainty associated with the imputations. The adjustment equation method always has the worst bias and coverage of the 95% confidence intervals. In some cases, the coverage of the 95% confidence interval for the adjustment equation is as low as 63% when estimating the mean (Table V) and 30% when estimating the prevalence (Table VI).

Table V. Estimates of the mean obtained under various models averaged over 1000 simulations $N = 2000$, $M = 25$.

$\pi =$ proportion missing	Scenario 1 under Bernoulli MCAR				Scenario 2 under Bernoulli MCAR				Scenario 1 under Systematic MCAR				Scenario 3 with square root transform				Scenario 3 with no transform			
	Estimate	Width 95% CI	CI Coverage (%)		Estimate	Width 95% CI	CI Coverage (%)		Estimate	Width 95% CI	CI Coverage (%)		Estimate	Width 95% CI	CI Coverage (%)		Estimate	Width 95% CI	CI Coverage (%)	
$\pi = 0.95$	Expected	60.00	NA	95.0	60.00	NA	95.0	60.00	60.00	NA	95.0	61.76	NA	95.0	61.76	NA	61.76	NA	95.0	95.0
	Full data	59.98	1.32	95.7	59.99	1.31	95.6	59.99	60.01	1.31	95.7	61.75	2.71	94.4	61.75	2.71	61.75	2.71	94.4	94.4
	complete case	59.98	5.94	95.0	60.06	5.94	95.1	60.06	60.00	5.96	100.0	61.86	12.31	94.5	61.86	12.31	61.86	12.31	94.5	94.5
	MI	60.01	2.79	95.7	60.01	5.56	96.1	60.01	60.01	2.77	94.5	61.85	5.83	95.1	61.81	5.84	61.81	5.84	94.4	94.4
	Stochastic CR	60.01	1.31	67.3	60.02	1.31	37.2	60.02	60.00	1.31	67.7	61.83	2.71	66.3	61.80	2.71	61.80	2.71	67.5	67.5
$\pi = 0.90$	AE	60.01	1.21	63.3	60.02	0.65	19.6	60.02	60.00	1.21	64.1	61.18	2.48	57.5	61.81	2.48	61.81	2.48	62.6	62.6
	Full data	60.02	1.31	94.8	60.00	1.32	93.5	60.00	60.00	1.31	94.9	61.74	2.71	95.1	61.74	2.71	61.74	2.71	95.1	95.1
	complete case	60.05	4.18	95.2	59.99	4.19	95.7	59.99	60.00	4.19	99.9	61.73	8.62	94.3	61.73	8.62	61.73	8.62	94.3	94.3
	MI	60.04	2.09	94.3	59.98	3.89	95.9	59.98	60.00	2.11	95.6	61.75	4.41	94.4	61.74	4.40	61.74	4.40	93.9	93.9
	Stochastic CR	60.04	1.31	77.9	59.97	1.32	50.9	59.97	60.00	1.31	78.7	61.74	2.71	74.4	61.73	2.71	61.73	2.71	73.3	73.3
$\pi = 0.85$	AE	60.04	1.21	75.4	59.97	0.66	27.2	59.97	60.01	1.21	76.4	61.10	2.48	65.0	61.74	2.48	61.74	2.48	71.0	71.0
	Full data	60.01	1.32	95.7	60.02	1.31	94.8	60.02	60.01	1.32	95.5	61.76	2.71	95.5	61.76	2.71	61.76	2.71	95.5	95.5
	complete case	59.97	3.41	94.2	60.06	3.40	95.4	60.06	60.03	3.23	100.0	61.69	7.02	94.2	61.69	7.02	61.69	7.02	94.2	94.2
	MI	59.99	1.85	93.5	60.06	3.16	94.3	60.06	60.03	1.79	94.6	61.76	3.85	95.4	61.75	3.85	61.75	3.85	95.6	95.6
	Stochastic CR	59.99	1.31	83.3	60.04	1.32	57.5	60.04	60.03	1.31	81.7	61.75	2.71	83.5	61.75	2.71	61.75	2.71	83.7	83.7
$\pi = 0.50$	AE	59.99	1.21	81.0	60.05	0.65	33.6	60.05	60.03	1.21	81.3	61.11	2.48	70.9	61.75	2.47	61.75	2.47	82.3	82.3
	Full data	59.99	1.31	94.9	60.01	1.32	94.3	60.01	59.99	1.32	95.7	61.77	2.71	94.0	61.77	2.71	61.77	2.71	94.0	94.0
	complete case	60.00	1.86	95.1	60.02	1.86	94.7	60.02	59.99	1.86	99.0	61.73	3.84	94.7	61.73	3.84	61.73	3.84	94.7	94.7
	MI	60.00	1.42	95.3	60.02	1.77	94.9	60.02	59.99	1.42	95.4	61.76	2.94	94.7	61.77	2.94	61.77	2.94	94.4	94.4
	Stochastic CR	60.00	1.31	94.0	60.01	1.32	80.9	60.01	59.99	1.32	93.2	61.77	2.71	91.5	61.77	2.71	61.77	2.71	92.2	92.2
	AE	60.00	1.21	90.8	60.02	0.66	56.3	60.02	59.99	1.21	91.1	61.13	2.48	77.4	61.76	2.48	61.76	2.48	90.0	90.0

Abbreviations: CI, Confidence Interval; NA, not applicable; MI, multiple imputation; Stochastic CR, stochastic conditional regression; AE, adjustment equation. MCAR, missing completely at random.

Table VI. Estimates of the prevalence of less than 30 obtained under various models averaged over 1000 simulations $N = 2000$, $M = 25$.

π = proportion missing	Scenario 1 under Bernoulli MCAR				Scenario 2 under Bernoulli MCAR				Scenario 1 under Systematic MCAR				Scenario 3 with square root transform				Scenario 3 with no transform			
	Estimate	Width 95% CI	CI Coverage (%)	Estimate	Width 95% CI	CI Coverage (%)	Estimate	Width 95% CI	CI Coverage (%)	Estimate	Width 95% CI	CI Coverage (%)	Estimate	Width 95% CI	CI Coverage (%)	Estimate	Width 95% CI	CI Coverage (%)	Estimate	Width 95% CI
Expected	2.23	NA	95.0	2.23	NA	95.0	2.23	NA	95.0	2.23	NA	95.0	14.43	NA	95.0	14.43	NA	95.0	14.43	NA
$\pi = 0.95$																				
Full data	2.29	1.31	95.7	2.28	1.30	94.2	2.26	1.30	93.6	14.40	3.08	95.3	14.40	3.08	95.3	14.40	3.08	95.3	14.40	3.08
complete case	2.26	5.70	99.8	2.13	5.57	99.9	2.36	5.81	100.0	14.28	13.62	93.5	14.28	13.62	93.5	14.28	13.62	93.5	14.28	13.62
MI	2.33	2.48	97.3	2.50	4.22	96.3	2.34	2.46	96.2	14.43	6.36	95.9	14.43	6.36	95.9	14.43	6.36	95.9	14.43	6.36
Stochastic CR	2.30	1.30	75.4	2.32	1.29	54.0	2.28	1.30	73.7	14.37	3.07	67.9	14.37	3.07	67.9	14.37	3.07	67.9	14.37	3.07
AE	1.50	1.05	29.7	0.02	0.19	0.0	1.49	1.05	29.1	12.32	2.87	32.9	11.57	2.79	21.6	11.57	2.79	21.6	11.57	2.79
$\pi = 0.90$																				
Full data	2.27	1.30	93.6	2.27	1.30	93.9	2.27	1.30	93.4	14.41	3.08	95.0	14.41	3.08	95.0	14.41	3.08	95.0	14.41	3.08
complete case	2.28	4.02	91.7	2.27	4.03	92.5	2.23	4.03	98.1	14.35	9.69	93.9	14.35	9.69	93.9	14.35	9.69	93.9	14.35	9.69
MI	2.28	2.01	97.4	2.43	3.05	96.9	2.30	2.01	97.7	14.43	5.01	97.1	14.43	5.01	97.1	14.43	5.01	97.1	14.43	5.01
Stochastic CR	2.26	1.30	81.6	2.36	1.31	62.2	2.29	1.30	83.1	14.42	3.08	79.0	14.42	3.08	79.0	14.42	3.08	79.0	14.42	3.08
AE	1.47	1.04	22.5	0.01	0.18	0.0	1.49	1.05	24.9	12.33	2.88	29.5	11.58	2.80	15.2	11.58	2.80	15.2	11.58	2.80
$\pi = 0.85$																				
Full data	2.28	1.30	93.9	2.27	1.30	93.5	2.26	1.30	92.2	14.43	3.08	95.1	14.43	3.08	95.1	14.43	3.08	95.1	14.43	3.08
complete case	2.29	3.33	91.9	2.27	3.31	90.7	2.23	3.14	97.5	14.49	7.95	95.6	14.49	7.95	95.6	14.49	7.95	95.6	14.49	7.95
MI	2.29	1.85	98.4	2.33	2.55	97.4	2.27	1.80	98.4	14.42	4.47	97.2	14.42	4.47	97.2	14.42	4.47	97.2	14.42	4.47
Stochastic CR	2.28	1.30	86.1	2.28	1.30	70.7	2.25	1.29	85.0	14.43	3.08	85.1	14.43	3.08	85.1	14.43	3.08	85.1	14.43	3.08
AE	1.49	1.05	24.9	0.01	0.18	0.0	1.47	1.04	23.4	12.32	2.88	27.2	11.54	2.80	11.5	11.54	2.80	11.5	11.54	2.80
$\pi = 0.50$																				
Full data	2.28	1.31	93.6	2.28	1.30	94.1	2.29	1.31	94.5	14.41	3.08	95.0	14.41	3.08	95.0	14.41	3.08	95.0	14.41	3.08
complete case	2.28	1.84	93.2	2.25	1.83	94.3	2.30	1.85	96.8	14.42	4.35	95.9	14.42	4.35	95.9	14.42	4.35	95.9	14.42	4.35
MI	2.28	1.50	97.8	2.28	1.70	98.8	2.30	1.51	97.2	14.41	3.50	97.9	14.45	3.65	99.0	14.45	3.65	99.0	14.45	3.65
Stochastic CR	2.27	1.30	90.6	2.27	1.30	89.6	2.30	1.31	93.5	14.40	3.08	93.1	14.45	3.08	93.6	14.45	3.08	93.6	14.45	3.08
AE	1.48	1.05	20.4	0.00	0.18	0.0	1.49	1.06	21.1	12.34	2.88	23.2	11.58	2.80	6.3	11.58	2.80	6.3	11.58	2.80

Abbreviations: CI, Confidence Interval; NA, not applicable; MI, multiple imputation; Stochastic CR, stochastic conditional regression; AE, adjustment equation. MCAR, missing completely at random.

Table VII. Estimates of the 2.5th percentile obtained under various models averaged over 1000 simulations $N = 2000$, $M = 25$.

$\pi =$ proportion missing	Scenario 1 under Bernoulli MCAR				Scenario 2 under Bernoulli MCAR				Scenario 1 under Systematic MCAR				Scenario 3 with square root transform				Scenario 3 with no transform			
	Estimate	Width 95% CI	CI Coverage (%)		Estimate	Width 95% CI	CI Coverage (%)		Estimate	Width 95% CI	CI Coverage (%)		Estimate	Width 95% CI	CI Coverage (%)		Estimate	Width 95% CI	CI Coverage (%)	
$\pi = 0.95$	Expected	30.60	NA	95.0	30.60	NA	95.0	30.60	30.60	NA	95.0	13.54	13.54	NA	95.0	95.0	13.54	NA	95.0	95.0
	Full data	30.63	3.60	95.5	30.65	3.62	94.4	30.70	30.70	3.63	95.2	13.65	13.65	3.51	94.9	94.9	13.65	3.51	94.9	94.9
	complete case	31.93	12.61	84.5	32.09	12.53	84.1	31.54	31.54	13.75	94.8	15.18	15.18	11.25	84.7	84.7	15.18	11.25	84.7	84.7
	MI	30.53	6.84	98.3	30.37	11.16	97.3	30.51	30.51	6.80	96.7	13.52	13.52	6.75	97.1	97.1	9.75	11.33	79.7	79.7
	Stochastic CR	30.68	3.61	76.6	30.76	3.63	55.5	30.72	30.72	3.62	75.8	13.70	13.70	3.49	74.8	74.8	10.06	4.89	23.6	23.6
$\pi = 0.90$	AE	33.09	3.32	37.0	45.50	1.78	0.0	33.11	33.11	3.32	37.6	16.21	16.21	3.52	36.0	36.0	17.34	3.36	21.5	21.5
	Full data	30.69	3.64	95.4	30.68	3.60	95.3	30.68	30.68	3.61	95.9	13.62	13.62	3.52	94.4	94.4	13.62	3.52	94.4	94.4
	complete case	31.31	15.26	93.7	31.26	15.34	93.3	31.31	31.31	16.37	99.3	14.46	14.46	12.52	93.1	93.1	14.46	12.52	93.1	93.1
	MI	30.60	5.60	98.1	30.35	8.20	97.0	30.53	30.53	5.61	98.2	13.54	13.54	5.55	98.1	98.1	10.02	8.70	68.5	68.5
	Stochastic CR	30.75	3.61	84.9	30.58	3.60	61.9	30.66	30.66	3.62	82.7	13.68	13.68	3.49	82.8	82.8	10.23	4.83	21.9	21.9
$\pi = 0.85$	AE	33.11	3.31	31.6	45.31	1.81	0.0	33.04	33.04	3.31	35.5	16.18	16.18	3.48	32.7	32.7	17.32	3.32	14.7	14.7
	Full data	30.66	3.61	94.2	30.70	3.63	94.5	30.69	30.69	3.58	94.6	13.67	13.67	3.51	95.4	95.4	13.67	3.51	95.4	95.4
	complete case	31.06	11.43	93.8	31.10	11.45	95.2	31.05	31.05	10.50	98.5	14.04	14.04	9.99	95.6	95.6	14.04	9.99	95.6	95.6
	MI	30.53	5.17	98.8	30.54	7.05	97.9	30.60	30.60	5.04	98.6	13.53	13.53	5.10	99.0	99.0	10.31	7.59	66.0	66.0
	Stochastic CR	30.68	3.61	87.5	30.73	3.65	71.1	30.73	30.73	3.58	87.1	13.68	13.68	3.56	88.5	88.5	10.57	4.72	24.6	24.6
$\pi = 0.50$	AE	33.04	3.36	35.2	45.44	1.80	0.0	33.11	33.11	3.31	34.1	16.18	16.18	3.51	32.2	32.2	17.39	3.34	10.3	10.3
	Full data	30.64	3.61	94.6	30.67	3.61	94.7	30.65	30.65	3.62	94.2	13.63	13.63	3.54	95.1	95.1	13.63	3.54	95.1	95.1
	complete case	30.73	5.23	93.5	30.78	5.19	95.5	30.70	30.70	5.26	96.2	13.63	13.63	5.04	94.1	94.1	13.63	5.04	94.1	94.1
	MI	30.55	4.28	98.4	30.55	4.81	99.2	30.51	30.51	4.29	98.0	13.48	13.48	4.24	98.8	98.8	11.70	5.22	80.7	80.7
	Stochastic CR	30.69	3.62	92.3	30.70	3.63	90.3	30.64	30.64	3.63	93.5	13.60	13.60	3.53	92.2	92.2	11.87	4.23	51.8	51.8
AE		33.06	3.33	32.1	45.39	1.80	0.0	33.01	33.01	3.33	33.0	16.10	16.10	3.48	29.3	29.3	17.30	3.31	5.1	5.1

Abbreviations: CI, Confidence Interval; NA, not applicable; MI, multiple imputation; Stochastic CR, stochastic conditional regression; AE, adjustment equation. MCAR, missing completely at random.

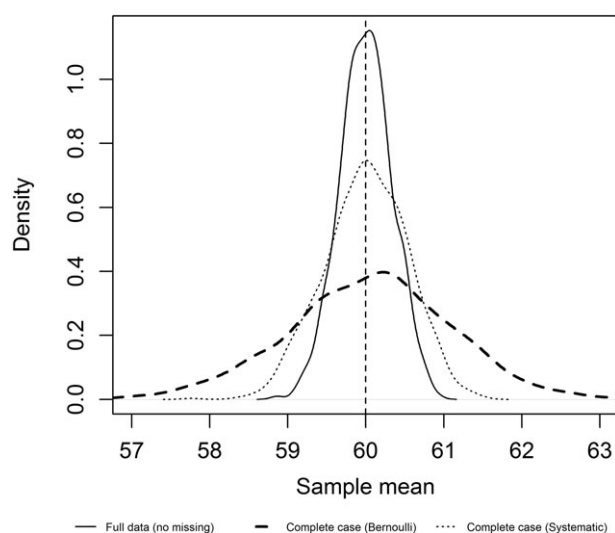


Figure 3. Sampling distribution of the sample mean based on 1000 simulations under scenario 1 with 90% missing data. The vertical line represents the true value of the parameter (population mean).

In contrast, multiple imputation produces unbiased estimates for all the estimands investigated and maintains approximately 95% coverage at all levels of missing data even when the correlation between X and Y is low (0.5). Figure 4 compares the sampling distribution of the 2.5th percentile under simulation scenario 1 (90% data missing). The adjustment equation's sampling distribution is shifted to the right of the other sampling distributions; demonstrating the severe bias in estimating features in the tails of the distribution.

Scenario 3 allows us to investigate the impact of ignoring the non-normality or pre-transforming the data to fit the imputation model. Under the conditions considered here (square root), the choice has limited impact when estimating the mean. However, it is clear that estimates at the tails are impacted, and it is preferable to use an imputation model that is faithful to the distribution of the original data. Figure 5 compares the sampling distribution of the 2.5th percentile (90% data missing) based on the adjustment equation, stochastic conditional regression and multiple imputation when the imputation model transforms the data or ignores the lack of normality. As with earlier simulations, the adjustment equation is severely biased when estimating features around the tails of the distribution, regardless of the model. On the other hand, both the multiple imputation and stochastic conditional regression models indicate that it is important to be loyal to the distribution of the data in order to obtain unbiased

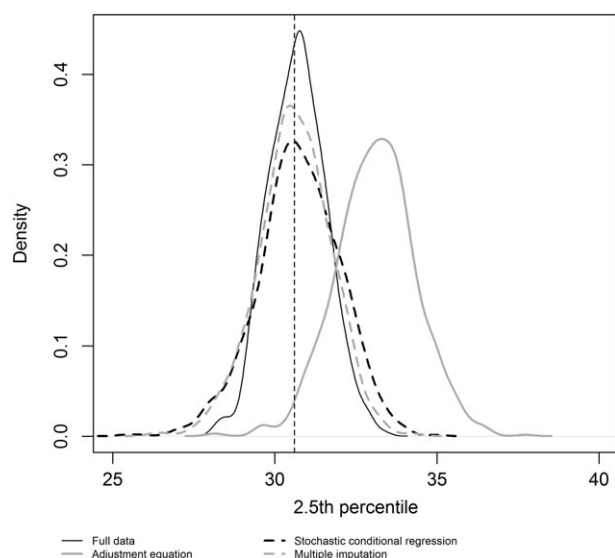


Figure 4. Sampling distribution of the 2.5th percentile based on 1000 simulations under scenario 1 with 90% missing data. The vertical line represents the true value of the parameter (population 2.5th percentile).

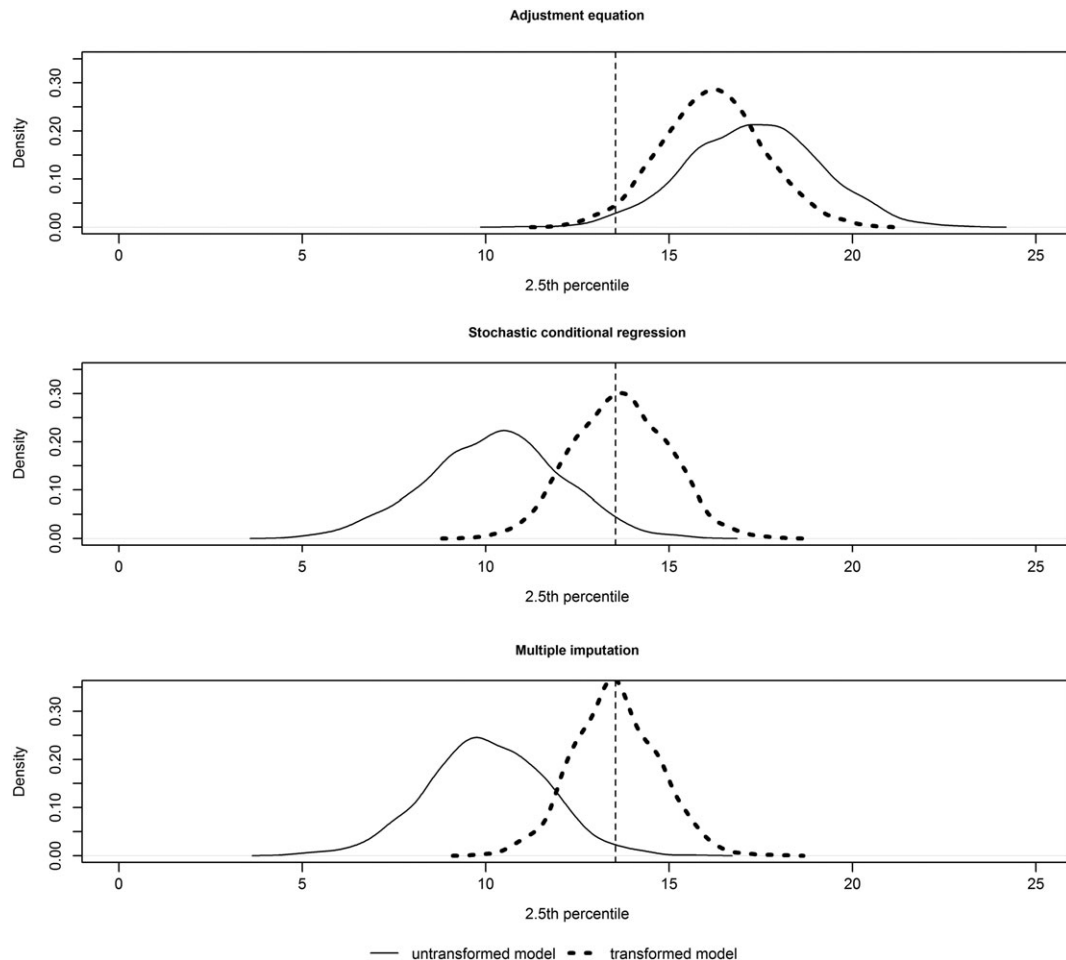


Figure 5. Sampling distribution of the 2.5th percentile based on 1000 simulations under scenario 3 with 90% missing data comparing models that either does or does not transform the data before fitting the imputation model. The vertical line represents the true value of the parameter (population 2.5th percentile).

estimates, especially for features around the tails. As with previous simulations, only the multiple imputation approach maintains at least 95% coverage for the confidence intervals; but even for multiple imputation, it can be important to select an imputation model that is consistent with the distribution of the original data.

In order to demonstrate how multiple imputation maintains the correlation between X and Y , the Pearson correlation was estimated using data from each method. While in the context of this setting, the correlation of X and Y may not be a parameter of direct interest, there are many other variables in a national survey whose correlation with the new assay may be of interest. If the correlation between X and Y is biased, then the correlation between any variable in the national survey and the newly imputed variable will be biased. Table S2 presents the average estimated correlation between X and Y for each method across 1000 simulations for each scenario. As expected, the correlation between the values imputed by the adjustment equation and X is exactly 1 with no sampling variability. Clearly, this is a result of using a linear transformation to impute the new assay from the old. The stochastic conditional regression approach appears to improve the bias of the correlation between X and Y but still underestimates the variability of the sample correlation leading to poor coverage of the 95% confidence intervals. Multiple imputation, in contrast, continues to achieve unbiased estimates, standard errors that properly reflect the uncertainty because of the imputation, and approximate 95% coverage.

6. Discussion

This paper describes an application of multiple imputation for nutritional biomarkers in large national surveys when there is an assay change over time and demonstrates that multiple imputation can be used

effectively to estimate various population parameters. This is important because estimands such as reference intervals (2.5th–97.5th percentile) are as important as central tendency, when describing nutritional biomarkers [22]. A key, untestable assumption for multiple imputation, is an ignorable missing data mechanism (i.e., MCAR or MAR). In our setting, the missingness is planned in advance and so the missing data mechanism is known. There is a tremendous amount of research and discussion about the performance of multiple imputation under far more complicated settings than the problem addressed in this paper [2–5,36,39,49–52]. For example, multiple imputation was used to bridge changes in the industry and occupation classification system used by the Bureau of Census [2–4] and to impute single-race categories from multiple-race reporting in the National Health Interview Survey [5]. Our proposal has several similarities to these papers: application to a large public-use dataset with access to many completely observed covariates; the need to impute a single variable to assess trends; a large percent of the data is missing; and the missing data mechanism is known. The previous work [2–5] imputed codes for categorical variables using logistic regression primarily with categorical covariates, where sparse data can be problematic. In contrast, the imputed variable (new assay) in our setting is continuous and based on a highly correlated continuous covariate (old assay, $r > 0.9$). This makes our setting nicely suited for multiple imputation under a multivariate normal distribution using regression methods for monotone missing data. Generally, in our setting, there is no reason to expect changes over time in the relationship between the two assays or with the other variables, as there might be with occupation and race codes. However, if there are assay drifts over time, as for the 25OHD RIA assay, the quality control measures at the CDC laboratory can detect them in advance and use this information as part of the study design, as was done for the 25OHD bridging study [16].

The simulations are intended to illustrate the improved properties of multiple imputation over the adjustment equation approach for selected estimands. However, the simulations may not be generalizable to complex surveys, as they are based on a simple random sample from infinite population. Previous research has shown that multiple imputation can yield biased estimates of variance when characteristics of the survey design are related to the variable(s) under study [53–55]. This is equally true of single imputation methods, so multiple imputation methods still offer more advantages than disadvantages in this regard. Simulations by Reiter *et al.* [55] suggest that ignoring the design variables in an imputation model may provide reasonable inferences if the outcome and the design variables have little correlation. These authors also estimated the bias of three estimates from NHANES by comparing the results from imputation models that either included or excluded the design variables. The results suggested that including design variables that are weakly associated with the outcomes led to more conservative inferences, unlike the current adjustment equation approach, which leads to liberal inferences. When analyzing the NHANES data in this paper, neither multiple imputation model considered included all the design variables. Therefore, there may be biases and/or inefficiencies that cannot be generalizable to other settings, and certainly not to other biomarkers. We showed empirically that including some of the design variables such as age, sex, race, and the survey weights had little practical effect on the resulting estimates, compared with a multiple imputation model with only using assay data, although on aggregate it did appear that the standard error was larger, suggesting that some of the covariates may be irrelevant. More research needs to be performed to determine the necessity and impact of including the design variables and other covariates into the proposed multiple imputation solution for this problem. An additional point, this paper primarily considered a single variable, the old assay, in developing the imputation models. This is because the method currently in use, the adjustment equation approach, generally only considers this variable (old assay). The primary purpose of this paper is demonstrate how multiple imputation can be performed with the same amount of observed information as the adjustment equation approach and provide much more valid inferences.

Using multiple imputation instead of the adjustment equation approach method with a properly designed bridging study does not require much more effort. Developing an imputation model under multiple imputation is equally challenging as under a single imputation method and subject to similar assumptions. Multiple imputation naturally incorporates missing data uncertainty leading to valid statistical inferences, unlike the adjustment equation method. In addition, the full Bayesian nature of the proposed solution can take advantage of the wealth of data and experience in the laboratory with the new and old assay, which can be used to inform the prior distributions of the Bayesian linear regression model. If necessary, one can build in measurement error for X , using historical information and a prior distribution. The selected functional form need not always be linear, as used in this paper. Rather one could include more flexible forms such as higher order terms or splines or utilize transformations. For example, a log transformation is useful with many nutritional biomarkers, as they

are often skewed right. Alternatively, more robust alternatives could be considered for the imputation model such as quantile regression [56]. The primary technical challenge with a multiple imputation approach is ensuring convergence [57,58].

Implementing the proposed solution does have its logistical challenges. All the data in NHANES are publically released on the internet allowing data analysts all over the world access to the data to obtain nationally representative estimates of the civilian US population. When there is an assay change, the online NHANES documentation is updated to provide the adjustment equation(s) developed from the bridging study [6–10], but no new data is released (re-release of vitamin D data was an exception). Currently, the documentation does not report the mean square error or the variance–covariance matrix of the coefficients of the adjustment equation. One recommendation is that the NHANES documentation publishes both the MSE and the variance–covariance matrix along with the estimated adjustment equation. This information could be used to either perform multiple imputation (see Appendix) [37] or at a minimum add a residual to the adjustment equation to ensure at least that estimates at the tails are less biased. However, those solutions are still undesirable because measurements made with the new assay are not used, because they are not publically available. We recommend that in addition to providing the MSE and variance–covariance matrix of the adjustment equation(s) in the documentation, the data is re-released to include the new assay values, where measured, along with the originally measured values. In this way, the data analyst can either use these data to develop their own multiple imputation model or use the published information in the documentation. The drawback is different multiple imputation models will produce slightly different national estimate(s). This is because the M multiple imputed values are simulated from the posterior conditional distribution of the missing data given the observed data; depending on the random number generator and the seed, every analyst will have an ever so slightly different estimate, even if the same imputation model and same statistical method is used. One way around this is to release M publically available data sets, where the agency develops the imputation model, as was done with the Dual X-ray data in NHANES [39,59]. There is an advantage to have the agency develop the multiple imputation model because they have access to far more information than the average data user including the original survey design variables and other confidential data. With this approach to the data release, data analysts will obtain the same estimates when using the M datasets for the same complete data statistical analysis. While this approach requires some added storage and some additional computational requirements to implement the multiple imputation solution, it is a minor inconvenience given software to do this is readily available in SAS, R, SPSS, Stata, S-Plus, and SUDAAN [23].

6.1. Disclaimer

The findings and conclusions in this report are those of the author(s) and do not necessarily represent the official position of the Centers for Disease Control and Prevention/the Agency for Toxic Substances and Disease Registry.

Appendix

Under a Bayesian formulation of simple linear regression model, we assume all the parameters, $\mu_x, \sigma_x^2, \beta_0, \beta_1, \sigma_{y|x}^2$, are random, and specify prior distributions. X is completely observed, therefore it is reasonable to assume that the parameters (μ_x, σ_x^2) are independent from $(\beta_0, \beta_1, \sigma_{y|x}^2)$ [32]. The bivariate normal likelihood can be written as the product of the marginal normal density of X and the conditional normal density of Y given X :

$$f(x_i, y_i | \mu, \Sigma) = f(x_i | \mu_x, \sigma_x^2) f(y_i | x_i, \beta_0, \beta_1, \sigma_{y|x}^2)$$

If there is no prior information, we can specify a non-informative Jeffrey's prior such that

$$P(\beta_0, \beta_1) \propto 1 \quad -\infty < \beta_0 < \infty, \quad -\infty < \beta_1 < \infty$$

so

$$P(\beta_0, \beta_1, \sigma_{y|x}^2) \propto \frac{1}{\sigma_{y|x}^2} \quad \sigma_{y|x}^2 > 0$$

With this assumption in place, the joint posterior distribution of $(\beta_0, \beta_1, \sigma_{y|x}^2)$ can be derived using the second term of the bivariate normal likelihood:

$$P(\beta_0, \beta_1, \sigma_{y|x}^2 | Y, X) \propto \prod_{i=1}^K f(y_i | x_i, \beta_0, \beta_1, \sigma_{y|x}^2) P(\beta_0, \beta_1, \sigma_{y|x}^2)$$

It can be shown from standard Bayesian theory with a Jeffrey's prior that the posterior distribution of $\sigma_{y|x}^2$ is $\sigma_{y|x}^2 | Y, X \sim \text{Inv} - \chi^2(K-2, \hat{\sigma}_{y|x}^2)$, where $\hat{\sigma}_{y|x}^2 = \frac{1}{K-2} \sum_{i=1}^K (y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i))^2$. The conditional posterior distribution of $\beta_0, \beta_1 | \sigma_{y|x}^2, X, Y$ is bivariate normal distribution centered around the least squares estimate for β_0, β_1 , such that

$$\beta_0, \beta_1 | \sigma_{y|x}^2, Y, X \sim \text{BVN} \left(\begin{bmatrix} \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} \\ \hat{\beta}_1 = \frac{\sum_{i=1}^K (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^K (x_i - \bar{x})^2} \end{bmatrix}, \begin{bmatrix} \frac{\sum_{i=1}^K x_i^2}{n \sum_{i=1}^K (x_i - \bar{x})^2} & \frac{-\sum_{i=1}^K x_i}{n \sum_{i=1}^K (x_i - \bar{x})^2} \\ \frac{-\sum_{i=1}^K x_i}{n \sum_{i=1}^K (x_i - \bar{x})^2} & \frac{1}{\sum_{i=1}^K (x_i - \bar{x})^2} \end{bmatrix} \sigma_{y|x}^2 \right).$$

These results suggest a way to create the multiple imputations simulated from the posterior predictive distribution of the parameters that accounts for the sampling variability in the linear regression, as well as uncertainty about the unknown model parameters $\beta_0, \beta_1, \sigma_{y|x}^2$.

Start with the available cases to get initial values $\hat{\beta}_0^{(0)}, \hat{\beta}_1^{(0)}, \hat{\sigma}_{y|x}^{2(0)}$ via least squares, followed by an iterative algorithm, known as data augmentation, that cycles between an imputation step (I-step) and posterior step (P-step). At the j th I-step we simulate the missing values from the conditional predictive distribution $Y_{miss,i}^{(j)} | x_i, Y_{obs}, \beta_0^{(j-1)}, \beta_1^{(j-1)}, \sigma_{y|x}^{2(j-1)}$. The P-step involves drawing from the conditional posterior distributions of $\hat{\sigma}_{y|x}^{2(j)}, \hat{\beta}_1^{(j)}, \hat{\beta}_0^{(j)} | Y_{miss}^{(j)}, Y_{obs}$, which under the Normal Bayesian linear model will have well-known probability distributions, to obtain updated estimates $\hat{\beta}_0^{(j)}, \hat{\beta}_1^{(j)}, \hat{\sigma}_{y|x}^{2(j)}$.

Specifically, we obtain a draw from $\sigma_{y|x}^{2(j)} | Y_{miss}^{(j)}, Y_{obs}$, by drawing from g , a chi-square random variable on $K-2$ degrees of freedom, and compute $\hat{\sigma}_{y|x}^{2(j)} = \frac{\hat{\sigma}_{y|x}^{2(0)}(K-2)}{g^{(j)}}$. Next, to draw from $\beta_0^{(j)}, \beta_1^{(j)} | \hat{\sigma}_{y|x}^{2(j)}, Y_{miss}^{(j)}, Y_{obs}, X$ we take two independent draws $z_0^{(j)}$ and $z_1^{(j)}$ from the standard normal distribution and compute

$$\begin{pmatrix} \beta_0^{(j)} \\ \beta_1^{(j)} \end{pmatrix} = \begin{pmatrix} \hat{\beta}_0^{(j-1)} \\ \hat{\beta}_1^{(j-1)} \end{pmatrix} + \hat{\sigma}_{y|x}^{2(j)} \begin{bmatrix} \sqrt{\frac{\sum_{i=1}^K x_i^2}{n \sum_{i=1}^K (x_i - \bar{x})^2}} & 0 \\ \frac{-\sum_{i=1}^K x_i}{\sqrt{\sum_{i=1}^K x_i^2} \sqrt{n \sum_{i=1}^K (x_i - \bar{x})^2}} & \frac{1}{\sqrt{\sum_{i=1}^K x_i^2}} \end{bmatrix} \begin{pmatrix} z_0^{(j)} \\ z_1^{(j)} \end{pmatrix}$$

Drawing from $\beta_0^{(j)}, \beta_1^{(j)} | \hat{\sigma}_{y|x}^{2(j)}, Y_{miss}^{(j)}, Y_{obs}, X$ can also be achieved by sampling from the conditional posterior distribution $\beta_1^{(j)} | \hat{\sigma}_{y|x}^{2(j)}, Y_{miss}^{(j)}, Y_{obs}, X \sim N\left(\hat{\beta}_1^{(j-1)}, \frac{\hat{\sigma}_{y|x}^{2(j)}}{K \sum_{i=1}^K (x_i - \bar{x})^2}\right)$ and then from the conditional posterior distribution of $\beta_0^{(j)}, \hat{\beta}_1^{(j)}, \hat{\sigma}_{y|x}^{2(j)}, Y_{miss}^{(j)}, Y_{obs}, X \sim N\left(\bar{y} - \hat{\beta}_1^{(j)} \bar{x}, \frac{\hat{\sigma}_{y|x}^{2(j)}}{K}\right)$.

Now, the i th subject's missing value can be drawn from the conditional predictive distribution $Y_{miss,i} | x_i, Y_{obs}, \beta_0, \beta_1, \sigma_{y|x}^2 \sim N\left(\beta_0 + \beta_1 x_i, \sigma_{y|x}^2\right)$ by computing $Y_{miss,i}^{(j+1)} = \hat{\beta}_0^{(j)} + \hat{\beta}_1^{(j)} x_i + r_i^{(j)} \hat{\sigma}_{y|x}^{(j)}$,

where r_i is drawn independently from standard normal distribution. In order to get M imputations for the i th subject with the new assay, these steps are repeated for $j=1 \dots M$.

Acknowledgments

I would like to thank Rayleen Lewis, Brody Ibeak, Dr. Rosemary Schleicher, Dr. Alula Hadgu and the reviewers for their careful reading of this article and helpful suggestions, which have led to a much improved version of the manuscript.

References

1. Graham JW, Hofer SM, MacKinnon DP. Maximizing the usefulness of data obtained with planned missing value patterns: an application of maximum likelihood procedures. *Multivariate Behavior Research* 1996; **31**(2):197–218.
2. Rubin D, Schenker N. Interval estimation from multiply-imputed data: a case study using census agriculture industry codes. *Journal of Official Statistics* 1987; **3**:375–387.
3. Schenker N, Treiman DJ, Weidman L. Analyses of public-use decennial census data with multiply-imputed industry and occupation codes. *Journal of the Royal Statistical Society, Series C (Applied Statistics)* 1993; **42**:545–556.
4. Clogg CC, Rubin DB, Schenker N, Schultz B, Weidman L. Multiple imputation of industry and occupation codes in census public-use samples using Bayesian logistic regression. *Journal of the American Statistical Association* 1991; **86**:68–78.
5. Schenker N, Parker J. From single-race reporting to multiple-race reporting: using imputation methods to bridge the transition. *Statistics in Medicine* 2003; **22**:1571–1587.
6. Curtin LR, Mohadjer LK, Dohrmann SM, Kruszan-Moran D, Mirel LB, Carroll MD, Hirsch R, Burt VL, Johnson CL. National Health and Nutrition Examination Survey: sample design, 2007–2010. Vital and Health Statistics, Series 2: Data Evaluation and Methods Research, 2013.
7. CDC/NCHS. February 3, 2014. Internet: https://www.cdc.gov/nchs/nhanes/about_nhanes.htm [accessed 23 August 2016].
8. CDC/NCHS. Internet: <https://wwwn.cdc.gov/nchs/nhanes/VitaminD/AnalyticalNote.aspx> [accessed 23 August 2016].
9. Centers for Disease Control and Prevention, National Center for Health Statistics. 2007–2008 Serum and red blood cell folate [cited 2011 Dec 17]. Available from: http://wwwn.cdc.gov/Nchs/Nhanes/2007-2008/FOLATE_E.htm [accessed 23 August 2016].
10. Centers for Disease Control and Prevention, National Center for Health Statistics. 2009–2010 Serum and red blood cell folate [cited 2011 Dec 17]. Available from: http://wwwn.cdc.gov/Nchs/Nhanes/2009-2010/FOLATE_F.htm [accessed 23 August 2016].
11. CDC/NCHS. Internet: http://wwwn.cdc.gov/nchs/nhanes/2003-2004/L06TFR_C.htm [accessed 23 August 2016].
12. CDC/NCHS. Internet: https://wwwn.cdc.gov/Nchs/Nhanes/2003-2004/L06MH_C.htm [accessed 23 August 2016].
13. CDC/NCHS. Internet: https://wwwn.cdc.gov/Nchs/Nhanes/2007-2008/GLU_E.htm [accessed 23 August 2016].
14. Tian L, Durazo-Arvizu RA, Myers G, Brooks S, Sarafin K, Sempos CT. The estimation of calibration equations for variables with heteroscedastic measurement errors. *Statistics in Medicine* 2014; **33**(25):4420–4436.
15. Cashman K, Kiely M, Kinsella M, Durazo-Arvizu R, Tian L, Zhang Y, Lucey A, Flynn A, Gibney MJ, Vesper HW, Phinney KW, Coates PM, Picciano MF, Sempos CT. Evaluation of vitamin D standardization program protocols for standardizing serum 25-hydroxyvitamin D data: a case study of the program's potential for national nutrition and health surveys. *American Journal of Clinical Nutrition* 2013; **97**:1235–1242.
16. Yetley EA, Pfeiffer CM, Schleicher RL, Phinney KW, Lacher DA, Christakos S, Eckfeldt JH, Fleet JC, Howard G, Hoofnagle AN, Hui SL, Lensmeyer GL, Massaro J, Peacock M, Rosner B, Wiebe D, Bailey RL, Coates PM, Looker AC, Sempos C, Johnson CL, Picciano MF, for the Vitamin D Roundtable on the NHANES Monitoring of Serum 25(OH)D: Assay Challenges Options for Resolving Them. NHANES monitoring of serum 25-hydroxyvitamin D: A roundtable summary. *Journal of Nutrition* 2010; **140**(11):2030S–2045S.
17. Pfeiffer CM, Zhang M, Lacher DA, Molloy AM, Tamura T, Yetley YA, Picciano MF, Johnson CL. Comparison of serum and red blood cell folate microbiologic assays for national population surveys. *Journal of Nutrition* 2011; **141**(7):1402–1409.
18. Pfeiffer CM, Hughes JP, Lacher DA, Bailey RL, Berry RJ, Zhang M, Yetley EA, Rader JJ, Sempos CT, Johnson CL. Estimation of trends in serum and RBC folate in the U.S. population from pre- to postfortification using assay-adjusted data from the NHANES 1988–2010. *Journal of Nutrition* 2012; **142**(5):886–893.

19. Schleicher RL, Sternberg MR, Lacher DA, Sempos CT, Looker AC, Durazo-Arvizu RA, Yetley EA, Chaudhary-Webb M, Maw KL, Pfeiffer CM, Johnson C. A method bridging study for serum 25-hydroxyvitamin D in NHANES 1988–2006: predicting standardized liquid chromatography-tandem mass spectrometry assay-equivalent concentrations from historical RIA data. *Natl Health Stat Report* 2016; **93**:1–16.
20. Schleicher RL, Sternberg MR, Lacher DA, Sempos CT, Looker AC, Durazo-Arvizu RA, Yetley EA, Chaudhary-Webb M, Maw KL, Pfeiffer CM, Johnson C. The vitamin D status of the U.S. population from 1988 to 2010 using standardized serum concentrations of 25-hydroxyvitamin D shows recent modest increases. *American Journal of Clinical Nutrition* 2016; **104**(2):454–461.
21. Pfeiffer CM, Osterloh JD, Kennedy-Stephenson J, Picciano MF, Yetley EA, Rader JI, Johnson CL. Trends in circulating concentrations of total homocysteine among US adolescents and adults: findings from the 1991–1994 and 1999–2004 National Health and Nutrition Examination Surveys. *Clinical Chemistry* 2008; **54**:801–813.
22. Centers for Disease Control and Prevention. National Center for Environmental Health. Second national report on biochemical indicators of diet and nutrition in the U.S. population 2012. Available at: <http://wwwn.cdc.gov/nutritionreport> [accessed 23 August 2016].
23. Yetley EA, Pfeiffer CM, Phinney KW, Fazili Z, Lacher DA, Bailey RL, Blackmore S, Bock JL, Brody LC, Carmel R, Curtin LR, Durazo-Arvizu RA, Eckfeldt JH, Green R, Gregory JF 3rd, Hoofnagle AN, Jacobsen DW, Jacques PF, Molloy AM, Massaro J, Mills JL, Nexo E, Rader JI, Selhub J, Sempos C, Shane B, Stabler S, Stover P, Tamura T, Tedstone A, Thorpe SJ, Coates PM, Johnson CL, Picciano MF. Biomarkers of folate status in NHANES: a roundtable summary. *American Journal of Clinical Nutrition* 2011; **94**(1):303S–312S.
24. Rubin DB. Multiple imputation after 18+ years (with discussion). *Journal of the American Statistical Association* 1996; **91**:473–489.
25. Graham JW. Missing data analysis: making it work in the real world. *Annual Review Psychology* 2009; **60**:549–576.
26. Harel O, Zhou X-H. Multiple imputation: review of theory, implementation and software. *Statistics in Medicine* 2007; **26**:3057–3077.
27. Horton NJ, Kleinman KP. Much ado about nothing: a comparison of missing data methods and software to fit incomplete data regression models. *The American Statistician* 2007; **61**:79–90.
28. Little RJA. Regression with missing X's: a review. *Journal of the American Statistical Association* 1992; **87**:1227–1237.
29. Schafer JL, Graham JW. Missing data: our view of the state of the art. *Psychological Methods* 2002; **7**:147–177.
30. Schafer JL. Multiple imputation: a primer. *Statistical Methods in Medical Research* 1999; **8**:3–15.
31. Schafer JL, Olsen MK. Multiple imputation for multivariate missing data problems: a data analyst's perspective. *Multivariate Behavior Research* 1998; **33**:545–571.
32. Rubin DB. Multiple Imputation for Nonresponse in Surveys, Wiley Series in Probability and Mathematical Statistics. Wiley: New York, US, 1987.
33. Little RJA, Rubin DB. Statistical Analysis with Missing Data (2nd edition). Wiley: Hoboken 2002.
34. Rao JNK, Shao J. Jackknife variance estimation with survey data under hot deck imputation. *Biometrika* 1992; **79**(4): 811–822.
35. Shao J, Sitter R. Bootstrap for imputed survey data. *Journal of the American Statistical Association* 1996; **93**:1278–1287.
36. Schafer JL, Ezatti-Rice TM, Johnson W, Khare M, Little RJA, Rubin DB. The NHANES III multiple imputation project. *Proceedings of the Section on Survey Research Methods of the American Statistical Association* 1996:28–37.
37. Raghunathan TE, Lepkowski JM, Hoewyk J, Solenberger PW. A multivariate technique for multiply imputing missing values using a sequence of regression models. *Survey Methodology* 2001; **27**:85–95.
38. van Buuren S, Brand J, Groothuis-Oudshoorn K, Rubin D. Fully conditional specification in multivariate imputation. *Journal of Statistical Computation and Simulation* 2006; **76**:1049–1064.
39. Schenker N, Borrud LG, Burt VL, Curtin LR, Flegal KM, Hughes J, Johnson CL, Looker AC, Mirel L. Multiple imputation of missing dual-energy X-ray absorptiometry data in the National Health and Nutrition Examination Survey. *Statistics in Medicine* 2011; **30**(3):260–276.
40. Barnard J, Rubin DB. Small-sample degrees of freedom with multiple imputation. *Biometrika* 1999; **86**:948–955.
41. SAS Institute Inc. Base SAS® 9.3 Procedures Guide. Cary, NC: SAS Institute Inc 2011.
42. SUDAAN. (Release 11.0.0 Build 308) Research Triangle. NC: RTI International, 2012.
43. Miller D. Reducing transformation bias in curve fitting. *American Statistician* 1984; **38**(2):124–126.
44. Institute of Medicine. Dietary Reference Intakes for Calcium and Vitamin D. National Academies Press: Washington, DC, 2010.
45. R Development Core Team R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria 2014. URL <http://www.R-project.org>.
46. van Buuren S, Groothuis-Oudshoorn K. MICE: Multivariate Imputation by Chain Equations in R. *Journal of Statistical Software* 2011; **45**(3):1–67.
47. Woodruff RS. Confidence intervals for medians and other position measures. *Journal of the American Statistical Association* 1952; **57**:622–627.
48. Lohr S. Sampling: Design and Analysis 2nd Edition. Brooks/Cole Publishing: Boston, US, 2010.
49. Barnard J, Meng X. Applications of multiple imputation in medical studies: from AIDS to NHANES. *Statistical Methods in Medical Research* 1999; **8**:17–36.
50. Lankester J, Perry S, Parsonnel J. Comparison of two methods – regression predictive model and intake shift model – for adjusting self-reported dietary recall of total energy intake of populations. *Frontiers of Public health* 2014; **2**(249):1–7.
51. Lee KJ, Carlin JB. Multiple imputation for missing data: fully conditional specification verses multivariate normal imputation. *American Journal of Epidemiology* 2010; **171**(5):624–632.
52. Sterne JAC, White IR, Carlin JB, Spratt M, Royston P, Kenward MG, Wood AM, Carpenter JR. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. *British Medical Journal* 2009; **338**:b2393.

53. Kim JK, Brick MJ, Fuller WA, Kalton G. On the bias of the multiple-imputation variance estimator in survey sampling. *Journal of the Royal Statistical Society B (Methodological)* 2006; **68**(3):509–521.
54. Kott PS. A paradox of multiple imputation. *Proc. Surv. Res. Meth. Sect. Am. Statist. Ass.* 1995:380–383.
55. Reiter JP, Raghunathan TE, Kinney SK. The importance of modeling the sampling design in multiple imputation for missing data. *Survey Methodology* 2006; **32**:143–149.
56. Wei Y, Ma Y, Carroll RJ. Multiple imputation in quantile regression. *Biometrika* 2012; **99**(2):428–438.
57. Abayomi K, Gelman A, Levy M. Diagnostics for multivariate imputations. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 2008; **57**:273–291.
58. White IR, Royston P, Wood AM. Multiple imputation using chained equations: issues and guidance for practice. *Statistics in Medicine* 2011; **30**(4):377–399.
59. CDC internet: https://wwwn.cdc.gov/Nchs/Data/Nhanes/Dxa/dxa_techdoc.pdf [accessed 23 August 2016].

Supporting information

Additional Supporting Information may be found online in the supporting information tab for this article.