Research Article

Statistics in Medicine

Received 12 August 2009, Accepted 4 August 2010 Published online 30 November 2010 in Wiley Online Library

(wileyonlinelibrary.com) DOI: 10.1002/sim.4080

Multiple imputation of missing dual-energy X-ray absorptiometry data in the National Health and Nutrition Examination Survey‡

Nathaniel Schenker*, †, Lori G. Borrud, Vicki L. Burt, Lester R. Curtin, Katherine M. Flegal, Jeffery Hughes, Clifford L. Johnson, Anne C. Looker and Lisa Mirel

In 1999, dual-energy x-ray absorptiometry (DXA) scans were added to the National Health and Nutrition Examination Survey (NHANES) to provide information on soft tissue composition and bone mineral content. However, in 1999–2004, DXA data were missing in whole or in part for about 21 per cent of the NHANES participants eligible for the DXA examination; and the missingness is associated with important characteristics such as body mass index and age. To handle this missing-data problem, multiple imputation of the missing DXA data was performed. Several features made the project interesting and challenging statistically, including the relationship between missingness on the DXA measures and the values of other variables; the highly multivariate nature of the variables being imputed; the need to transform the DXA variables during the imputation process; the desire to use a large number of non-DXA predictors, many of which had small amounts of missing data themselves, in the imputation models; the use of lower bounds in the imputation procedure; and relationships between the DXA variables and other variables, which helped both in creating and evaluating the imputations. This paper describes the imputation models, methods, and evaluations for this publicly available data resource and demonstrates properties of the imputations via examples of analyses of the data. The analyses suggest that imputation helps to correct biases that occur in estimates based on the data without imputation, and that it helps to increase the precision of estimates as well. Moreover, multiple imputation usually yields larger estimated standard errors than those obtained with single imputation. Published in 2010 by John Wiley & Sons, Ltd.

Keywords: body composition; body fat; body mass index; bone mineral density; missing at random; sequential regression multivariate imputation

1. Overview

Dual-energy x-ray absorptiometry (DXA) is used to measure bone mineral density [1–4] and has become one of the most widely accepted methods of measuring soft tissue composition as well, due to its speed, ease of use, and strong correlation with criterion methods of assessing body composition [5, 6]. In 1999, whole-body DXA scans were added to the National Health and Nutrition Examination Survey (NHANES), a nationally representative sample survey of the civilian non-institutionalized population in the U.S., to provide information on soft tissue composition and bone mineral content. However, in the survey years 1999–2004, DXA data were missing in whole or in part for about 21 per cent of the NHANES participants eligible for the DXA examination. Moreover, the missingness is associated with important characteristics such as body mass index (BMI) and age. BMI ((weight in kilograms)/(height in meters)^2) is used as the standard measure of obesity in the U.S. and other countries [7].

In general, when missingness is related to characteristics of the individuals surveyed, that is, the missing data are not 'missing completely at random (MCAR)' [8], analyzing only the cases with non-missing data is known to produce biased estimates [9]. Techniques such as multiple imputation help to reduce such biases by adjusting for differences between the complete cases and incomplete cases on variables that are observed. Moreover, by incorporating partially

National Center for Health Statistics, Centers for Disease Control and Prevention, Hyattsville, MD, U.S.A.
**Correspondence to: Nathaniel Schenker, National Center for Health Statistics, Centers for Disease Control and Prevention, 3311 Toledo Road, Room 3209, Hyattsville, MD 20782, U.S.A.*
†E-mail: nschenker@cdc.gov

observed data for incomplete cases as well as information contained in other variables used for prediction, imputation helps to increase the precision of estimates relative to analyzing the unimputed data. The reduction of bias and increase in precision tend to be magnified when the amount of missing data is large. Finally, imputation creates a completed data set, so that standard analyses designed for complete data can be applied.

When a single value has been imputed for each missing datum, and then analyses are conducted that treat the imputed values as if they were true values, the uncertainty in the imputations is not reflected in the analyses. Thus, for example, estimated standard errors based on singly imputed data tend to be too small. In general, the bias of an estimated standard error under single imputation tends to increase as the amount of missing data increases and tends to decrease when the predictors in the imputation model are stronger. The bias also depends on what is being estimated, however, and it is difficult to assess in advance for any specific case. Multiple imputation [10–12] is a method designed to allow analyses to reflect the extra uncertainty due to missing data. With multiple imputation, several, say M, sets of imputations for the missing values are drawn from an approximate posterior predictive distribution conditional on the observed values, creating M completed data sets. Each completed data set is analyzed, and then the M analyses are combined via simple formulas [10–12]. Many statistical packages have recently developed routines for creating and/or analyzing multiply imputed data sets under selected models [12].

To facilitate analyses of all of the observed NHANES DXA data, adjust for differential missingness, incorporate information from other available NHANES data, and allow analyses to reflect the extra uncertainty due to missing data, multiple imputation of the missing DXA data was performed for the 1999–2004 survey years. The multiply imputed DXA data are publicly available at http://www.cdc.gov/nchs/nhanes/dxx/dxa.htm (accessed 27 May 2010). Work is currently underway to multiply impute missing DXA data for the 2005 and 2006 survey years.

Several features made the multiple-imputation project interesting and challenging statistically. These included the relationship between missingness on the DXA measures and the values of other variables, such as BMI and age; the highly multivariate nature of the main variables of interest (32 DXA measures), with high correlations among them; the need to transform the DXA variables during the imputation process, and the resulting use of an iterative Box–Cox [13] estimation procedure; the desire to use a large number of non-DXA predictors, many of which had small amounts of missing data themselves, in the imputation models; the use of lower bounds in the imputation procedure; and relationships between the DXA variables and other measured variables, which helped in both predicting the missing values and evaluating the imputed values.

This paper discusses various aspects of the multiple-imputation project and thereby gives several examples of the types of considerations involved in creating imputations for a large, publicly available, data set. Section 2 briefly describes the NHANES sample design, the DXA data, and the missingness of the DXA data. Section 3 describes the multiply imputed data. Sections 4 and 5 discuss the models, methods, and evaluations used in producing the multiple imputations. Section 6 demonstrates properties of the imputations via examples of analyses of the data. The paper concludes with a discussion in Section 7.

2. The NHANES sample design and DXA data

2.1. The sample design

The complex sample design of the NHANES is described briefly here. Further details are available in other publications [14, 15] and in the NHANES Web Tutorial (http://www.cdc.gov/nchs/tutorials/Nhanes/index_current.htm, accessed 27 May 2010). The NHANES uses a four-stage sample design. The first stage involves selection of primary sampling units (PSUs), which are usually counties but are occasionally combinations of counties. Factors used in the selection of PSUs include geography, metropolitan-area status, contextual information on racial composition and the percentage of the population below the poverty level, and a measure of size. The second stage samples area segments within each sampled PSU. The segments are census blocks or combinations of blocks. In the third stage, households and non-institutional group quarters (e.g. dormitories) are sampled within each sampled segment and screened to identify potential sampled persons. The fourth stage samples persons within the screened dwellings, taking into account age, sex, race/ethnicity, and income information obtained during screening. In 1999–2004, the NHANES sampling fractions were set to 'oversample' Mexican Americans, non-Hispanic black Americans, low-income non-Hispanic white and other Americans, adolescents of ages 12–19, and persons above age 60.

The NHANES obtains data from interviews in respondents' homes, followed by physical examinations in mobile examination centers. For 1999–2004, data were released publicly in three two-year cycles (1999–2000, 2001–2002, and 2003–2004). To allow analysts to account for the sample design, the public-use files contain weights, which account for the differential probabilities of selection, non-response, and non-coverage. Because not all interviewed persons received physical examinations, different weights are provided for analyses involving data from the interviews alone versus analyses

involving data from the physical examinations. To protect confidentiality while allowing analysts to obtain approximately valid variance estimates, the public-use files contain additional 'pseudo' design variables, in the form of strata and masked variance units, which can be treated as strata and PSUs with standard software for analyzing complex survey data. The public-use strata and masked variance units do not correspond directly to features of the actual sample design, but are intended to yield variance estimates close to those that would be obtained with the actual design features [16].

2.2. The DXA data and their missingness

In 1999–2004, DXA survey participants 8 years of age and older who came to the mobile examination centers were eligible for the DXA examination, except pregnant females (based on self-report or a laboratory test). Eligible individuals were excluded if they reported recent tests with radiographic contrast material, recent participation in nuclear medicine studies, or weights or heights over the DXA table limit (300 pounds or 6 feet, 5 inches). In 1999, females 8–17 years of age were also excluded due to concerns over how to handle the reporting of pregnancy test results for minors.

The whole-body DXA scans were acquired using a Hologic QDR 4500A fan-beam densitometer (Hologic, Inc., Bedford, MA). They were reviewed and analyzed by staff in the Department of Radiology, University of California, San Francisco. Further details of the protocols are available [17]. Soft tissue and bone measures were obtained for the total body and body regions.

Of the over 22 000 persons eligible for the DXA examination in 1999–2004, about 21 per cent had one or more missing DXA values. Reasons for missingness, in addition to the exclusions mentioned earlier, included the presence in scans of non-removable objects (prostheses, pacemakers, implants, and casts); excessive x-ray 'noise' due to truncal adiposity, which interfered with the ability to obtain valid measurements; arm/leg overlap; body parts out of the DXA table scan area; positioning problems (head, arms/hands, or feet turned); and other factors. The percentages with data missing for one or more body regions are provided in Table I by age and in Table II by BMI. The percentages tend to increase with increasing age, due in part to the higher prevalence of non-removable objects in scans of older people, and with BMI, due in part to truncal adiposity 'noise' and exclusions of persons with high weights.

As implied by the general discussion in Section 1, because DXA data missingness is related to age, BMI, weight and height (due in part to exclusions), and other characteristics, participants with missing data cannot be treated as a completely random subset of the original sample. Otherwise, analytic results may be biased toward participants with lower amounts of missing data [9]. Of particular concern is the systematic pattern of missingness by BMI, given BMI's relation to body fatness.

Table I. Percentages of persons with data missing for one or more DXA regions, by age, NHANES, 1999–2004.

Age (years)	Percentage with missing data
8–11	15
12–15	17
16–19	16
20–29	19
30–39	21
40–49	22
50–59	24
60–69	26
70–79	30
80+	42

Table II. Percentages of persons with data missing for one or more DXA regions, by body mass index*, NHANES, 1999–2004.

Body mass index	Percentage with missing data
<20	15
20 – <25	16
25 – <30	16
30 – <35	20
35 – <40	38
⩾40	75

*The calculations for this table did not include persons (2.8 per cent) for whom BMI is missing.

Published in 2010 by John Wiley & Sons, Ltd.
Statist. Med. ***2011***, 30 260–276

3. The multiply imputed DXA data

Multiple imputation of missing DXA data in the NHANES was performed for all eligible (i.e. non-pregnant) participants except those with amputations other than fingers or toes. As has often been done with public-use data [18–22], a small number of imputations of the missing values was created ($M=5$) to limit burden for analysts. (See Section 7 for further discussion on the number of imputations.) Accordingly, five sets of completed DXA data were created. The imputed DXA values differ across the five completed data sets, whereas the non-missing DXA values remain the same (no imputation was performed for them).

The multiply imputed data files, with documentation, can be found at http://www.cdc.gov/nchs/nhanes/dxx/dxa.htm (accessed 27 May 2010). Variables in the files include fat mass, body fat percentage, fat-free mass (includes bone mineral content), lean soft tissue mass (excludes bone mineral content), bone area, bone mineral content, and bone mineral density for the total body and several body subregions depending on the variable being considered. Also included are indicators of which DXA values were imputed and reasons for missingness. Because NHANES data are released in two-year cycles, the imputed DXA data for the aforementioned females 8–17 years of age in 1999, as well as the observed and imputed DXA data for females of the same ages in 2000, were not included in the public-use data release for reasons of confidentiality. These data can be accessed through the National Center for Health Statistics Research Data Center (http://www.cdc.gov/rdc/, accessed 27 May 2010).

Since multiple imputation of a variable reflects uncertainty due to missing data, some variability among the five imputed values for a participant is to be expected. However, for the participants (roughly 400 for 1999–2004) who were missing measured weight and waist circumference (important predictors for imputation) and also the values of all of their DXA variables, the imputed DXA values were found to vary extremely (from very low values to very high values) among the five imputations. Because of this extreme variability, the imputed values for these participants have been placed in separate files (http://www.cdc.gov/nchs/nhanes/dxx/dxa_highlyvariable.htm, accessed 27 May 2010). Analysts should be aware of the highly variable nature of the imputed values when considering the use of these separate files. Theoretically, the extreme variability would be expected, due to the lack of information in the observed data on which to condition the imputations. Given this lack of information, the predictive distribution of the missing values tends to have high dispersion, and thus the imputations vary widely. Omitting persons with such highly variable imputations could under-represent the uncertainty in the imputed values and thus result in downward biases in variance estimates. Moreover, if the missingness of data for such persons is related to the outcome of interest, omitting the persons could result in biases in point estimates. Even if the persons were not omitted, however, it would be difficult to adjust for any biases in point estimates because of the missing data in the key auxiliary variables. Fortunately, the number of persons in question is small (about 2 per cent of all persons eligible for the DXA examination and 8 per cent of persons with missing DXA values); hence, the effects of the highly variable imputations are limited.

4. Models and methods for multiply imputing the NHANES DXA data

The development and evaluation of the imputation procedures were performed iteratively, with a given model leading to a set of evaluation results, leading to modifications of the model, and so on. Features of the final model and methods will be described first, followed by a description in Section 5 of evaluations performed.

4.1. Sequential regression multivariate imputation

Multiple imputations were created using sequential regression multivariate imputation (SRMI) [23], as implemented by the module IMPUTE in the software IVEware [24]. Other computations during the imputation process were performed using SAS [25], which was also used to call IVEware.

A brief description of SRMI is as follows. Let X denote the fully observed variables, and let $Y_1, Y_2, \ldots, Y_k$ denote the variables with missing values. The imputation process for $Y_1, Y_2, \ldots, Y_k$ proceeds in c iterations. In each iteration, the algorithm cycles through $Y_1, Y_2, \ldots, Y_k$, imputing the missing values of each variable, say Y_j, randomly from an approximate predictive distribution for Y_j given all of the other variables $(X, Y_1, Y_2, \ldots, Y_{j-1}, Y_{j+1}, Y_{j+2}, \ldots, Y_k)$. The approximate predictive distribution is based on the regression of Y_j on the other variables, fitted to the cases with Y_j not missing. When the regression is fitted, any missing values of the other variables are set equal to their most recently imputed values. After c iterations, the final imputations of the missing values in $Y_1, Y_2, \ldots, Y_k$ are used.

By including all variables other than the variable being imputed as predictors in each regression model, and by iteratively cycling through the various regression models, SRMI builds in relationships among the variables. This was especially important for the DXA variables, which are highly interrelated. Moreover, for each regression in the SRMI procedure, IVEware provides a choice of models, depending on the form of the outcome variable (continuous, categorical,

Table III. DXA variables included in the imputation model, with percentages of data missing, NHANES, 1999–2004.

Variable	Missing data (per cent)
Left arm area (cm^2)	13.8
Left arm bone mineral content (g)	13.8
Left arm fat (g)	14.1
Left arm lean excluding bone mineral content (g)	14.1
Right arm area (cm^2)	14.0
Right arm bone mineral content (g)	14.0
Right arm fat (g)	14.4
Right arm lean excluding bone mineral content (g)	14.4
Trunk fat (g)	17.5
Trunk lean excluding bone mineral content (g)	17.5
Left leg area (cm^2)	14.7
Left leg bone mineral content (g)	14.7
Left Leg Fat (g)	15.0
Left Leg Lean excluding Bone Mineral Content (g)	15.0
Right leg area (cm^2)	14.9
Right leg bone mineral content (g)	14.9
Right leg fat (g)	15.3
Right leg lean excluding bone mineral content (g)	15.3
Head area (cm^2)	12.7
Head bone mineral content (g)	12.7
Head fat (g)	12.7
Head lean excluding bone mineral content (g)	12.7
Left ribs area (cm^2)	17.5
Left ribs bone mineral content (g)	17.5
Right ribs area (cm^2)	17.5
Right ribs bone mineral content (g)	17.5
Thoracic spine area (cm^2)	17.5
Thoracic spine bone mineral content (g)	17.5
Lumbar spine area (cm^2)	17.5
Lumbar spine bone mineral content (g)	17.5
Pelvis area (cm^2)	17.5
Pelvis bone mineral content (g)	17.5

count, or semi-continuous), and it allows bounds to be placed on the variables being imputed, among other features. The modeling flexibility and allowance of bounds were particularly helpful for multiply imputing NHANES DXA data, for two reasons. First, as described in more detail in Section 4.2.2, continuous and categorical non-DXA variables were included in the process to help predict the missing DXA values, and some non-DXA variables had missing values as well. This resulted in the need to impute both continuous and categorical variables during the process. (A small number of the categorical non-DXA variables could be regarded as ordinal. In the imputation process, the ordering of the categories was ignored because IVEware does not provide special models for ordinal variables.) Second, bounds were placed on some of the imputed values, as described in Section 4.2.4.

Non-DXA variables were included in the process solely to help predict the missing DXA values. The imputations for missing values of the non-DXA variables were not included in the public data release, because the focus of the project was on handling missing data in the DXA variables. Moreover, the missingness rates for the non-DXA variables were generally much lower than those for the DXA variables (as demonstrated in Tables III and IV), and including imputations for the non-DXA variables in the public data release would have resulted in a very large and complex data file, which might have discouraged analysts from using the DXA data. However, because the missing non-DXA values were multiply imputed as part of the process, the uncertainty in those missing values was reflected in the multiple imputations of the missing DXA values.

Unlike traditional Bayesian iterative simulation methods for handling missing data (e.g. [26]), SRMI does not require the specification of a joint distribution for all of the Y-variables given X, but rather only the specification of individual regression models for each of the Y-variables. Therefore, it does not necessarily imply a joint model for all of the Y-variables conditional on X [23, 27]. However, the flexibility of the procedure regarding various types of variables and placing bounds on the variables being imputed, as well as its ability to handle a large number of variables with missing values, allowed it to handle complicating factors in the NHANES DXA imputation project that would have been very difficult to handle using a method based on a full joint model. Moreover, had all of the variables with missing values been continuous (i.e. had there been no missing values among the categorical non-DXA variables) and had no

Published in 2010 by John Wiley & Sons, Ltd.
Statist. Med. ***2011***, 30 260–276

Table IV. Non-DXA variables included in the imputation model, with percentages of data missing, NHANES, 1999–2004.

Variable	Age range (years)	Missing data (per cent)
Self-reported race/ethnicity (four categories)	8+	0
Age at screening	8+	0
Annual income (two categories)	8+	4.1
Education (three categories)	20+	0.2
Metropolitan area? (yes/no)	8+	0
Region (four categories)	8+	0
2-year mobile examination center weight	8+	0
Data release cycle (three 2-year cycles)	8+	0
Average diastolic blood pressure (mmHg)	8+	5.5
Average systolic blood pressure (mmHg)	8+	4.7
Taking prescribed medicine for blood pressure? (yes/no)	16+	0.3
Weight (kg)	8+	2.2
Arm circumference (cm)	8+	3.2
Subscapular skinfold (mm)	8+	16.8
Triceps skinfold (mm)	8+	8.9
Waist circumference (cm)	8+	4.1
Height (cm)	8+	1.8
Waist circumference (cm)/height (cm)	8+	4.5
Body mass index (kg/m^2)	8+	2.8
Body mass index level (three categories)	20+	3.5
Doctor said you have diabetes? (yes/no)	8+	0.1
General health condition (three categories)	8+	0.1
Total cholesterol (mg/dL)	8+	8.2
HDL cholesterol (mg/dL)	8+	8.2
Natural log trigylcerides (mg/dL) non-fasting	12+	7.5
Taking prescribed medicine for cholesterol? (yes/no)	20+	4.2
Activity compared to others the same age (two categories)	12+	2.1
Moderate or vigorous activity in past 30 days? (yes/no)	12+	1.1
Smoking status (three categories)	12+	2.9
Experienced a fracture? (yes/no)	20+	0.4
Family history of osteoporosis? (yes/no)	20+	3.9
Ever used birth control (females only; yes/no)	12+	10.5
Menopause status (females only; four categories)	12+	12.8
Ever took estrogen for other than non-birth control? (females only; yes/no)	20+	11.9
Took anticonvulsants in past 30 days? (yes/no)	8+	0.1
Took thyroid hormones in past 30 days? (yes/no)	8+	0.1
Took cortisone glucocoticoids in past 30 days? (yes/no)	8+	0.1
Average alcohol consumption per year	20+	8.0
Took calcium or vitamin D supplements in past 30 days? (yes/no)	8+	0.2
Took antacids in past 30 days? (yes/no)	8+	0.2
Milk consumption in past 30 days (two categories)	8+	0.1
Dietary intake yesterday (11 items)	8+	5.4

bounds been placed on the imputations, the SRMI-based imputation procedure would actually have been equivalent to imputation based on a multivariate normal model (conditional on transformations of the DXA variables, discussed in the next section), with the SRMI steps corresponding to the conditional posterior distributions in the Gibbs sampling [23]. Finally, methods such as SRMI typically have been found to work well in practice (e.g. [27, 28]).

4.2. Further details of the imputation model

4.2.1. Imputation within gender and age groups. The distributions of the DXA and non-DXA variables vary by gender and age, as do the availability of some of the non-DXA variables, the missing-data rates and reasons for missingness, and perhaps the mechanisms relating the variables to each other. Therefore, the imputation procedure was implemented separately within 10 gender by age (8–11, 12–19, 20–39, 40–59, or 60+ years) groups, with age also retained as a predictor to account for variability within each group. The selected gender-by-age groupings are also commonly used in analyses of the NHANES data.

Ten iterations of the SRMI procedure were used to create one completed data set, as suggested by Raghunathan *et al.* [24]; results were not sensitive to the number of iterations. To create multiple imputations, the procedure was repeated independently 5 times. After imputation, the data for the gender-age groups were concatenated.

4.2.2. Selection of variables. It is beneficial to include a large number of predictors in models for multiple imputation, especially variables that are predictive of the items being imputed, variables related to missingness, and variables that will be used in subsequent analyses of the multiply imputed data [29–32]. Therefore, many such predictors were included in the imputation models for the DXA data in addition to the DXA variables themselves; the predictors included demographic, socioeconomic, and geographic variables, body measurements, health indicators, dietary and medication use variables, and blood test results. A categorical variable for the data release cycle (1999–2000, 2001–2002, 2003–2004) was also included to account for possible changes over time. Complete lists of the DXA and non-DXA variables in the models, together with missing-data percentages, are given in Tables III and IV.

Rubin [30] and Reiter *et al.* [32] discussed the importance of including variables reflecting the sample design in the imputation model. To reflect the NHANES sample design in the DXA imputation model, the mobile examination center weights were included, as were several variables related to the selection of PSUs: metropolitan-area status, data release cycle, self-reported race/ethnicity, family income, and education. Moreover, several strong person-level predictors for the DXA variables, such as other DXA variables, age, race/ethnicity, gender, BMI, waist circumference, triceps and subscapular skinfolds, report of previous fracture, and family history of osteoporosis were included. Thus, any residual associations between the sample design features and the DXA variables should be small.

4.2.3. Transformations of the DXA variables. The linear regression models for the DXA variables in the SRMI procedure assume normal distributions for the error terms. To make this assumption more tenable, transformations were used for the DXA variables in the imputation procedure. After multiple imputations were created on the transformed scales, the imputed values were back-transformed to their original scales.

Several alternative approaches to transforming the DXA variables were explored, including no transformation, logarithmic transformation, and iteratively estimated Box–Cox [13] power transformations. When no transformation was used, the imputation procedure tended to produce physiologically implausible minimum values. When the logarithmic transformation was used for all of the DXA variables, imputed values at the upper ends of distributions appeared overinflated. After additional exploration of alternatives, the iterative Box–Cox method was selected. This method, which was implemented separately within each of the 10 gender-age groups used for imputation, estimated transformations based on the complete cases, prior to imputation. The estimated transformations were then applied to the entire data set for the imputation process. Thus, the uncertainty due to estimating the transformations was not incorporated into the imputations.

During each iteration, the following steps were performed. First, the 32 DXA variables were ordered randomly. The method then cycled through the DXA variables, updating the estimated transformation for each variable before proceeding to the next variable. The transformation for each variable was estimated via maximum likelihood estimation for the normal linear regression model [33], with the predictors that would be used to impute missing values of that variable in the SRMI procedure. Maximum likelihood estimation was implemented via the SAS procedure TRANSREG, which by default considered values between -3 and 3, in increments of 0.25, for the possible exponents in the power transformations.

The iterations continued until the set of estimated transformations stabilized, that is, did not change from one iteration to the next. The exponents for the final estimated power transformations of the DXA variables ranged from -0.5 to 2.5.

4.2.4. Setting lower bounds. Although transformed DXA values were used in the imputation process, occasionally there were implausibly low imputed values. This issue was resolved by placing lower bounds on the imputed values, as allowed by IVEware. Within each of the gender-age groups used for imputation, a lower bound for each DXA variable (before transformation) was set at the minimum of the non-missing values divided by $\sqrt{2}$. Other possible lower bounds had been tried as well. For example, using a lower bound of 0 still resulted in some implausibly low imputed values (although it at least ensured that the imputed values would be positive). Using the minimum of the non-missing values as the lower bound resulted in compression of the distribution of imputed values at the lower end as well as an upward shift of the low percentiles of the distribution. Although the use of the minimum of the non-missing values divided by $\sqrt{2}$ was somewhat arbitrary, it resulted in reasonable imputations. The imputed values at both the low and high ends of the distribution were judged to be plausible, and there was not the type of compression at the low end that occurred with the use of the minimum of the non-missing values itself. Moreover, there was some precedent for the use of the minimum of the non-missing values divided by $\sqrt{2}$ with the NHANES laboratory data. Laboratory values that are below the level of detection (LOD) are calculated as LOD divided by $\sqrt{2}$ [34].

As discussed in Section 4.1, the non-DXA variables were not the focus of this project and had lower amounts of missingness than the DXA variables; and imputed values for them were not included in the public data release. Nevertheless, lower bounds of 0 were used for the non-negative, non-categorical variables (with 0 included or excluded as appropriate) during the imputation process, to ensure that their imputed values would be in the range of possible values.

Published in 2010 by John Wiley & Sons, Ltd.
Statist. Med. ***2011***, 30 260–276

4.3. The assumption of missingness at random

The imputation methods described in Sections 4.1 and 4.2 involve an implicit assumption that the missing data are 'missing at random' (MAR) [8], which means heuristically that conditional upon quantities that are observed and included in the imputation model, the probabilities of missingness do not depend on the missing values. In other words, the probabilities of missingness depend directly only on the observed quantities that are in the model. MAR is a much less stringent assumption than MCAR, which means heuristically that the probabilities of missingness do not depend at all on the characteristics of the individuals surveyed, or equivalently that the individuals with missing values are a simple random sample of all the individuals in the survey. MCAR clearly does not hold for the DXA imputation project; see, e.g. Sections 1 and 2.2. An even less stringent assumption than MAR would be that the missing data are 'missing not at random' (MNAR), that is, that the probabilities of missingness depend on the missing values even after conditioning on the observed quantities in the model.

In practice, most imputation procedures assume MAR rather than MNAR [35, 36]. There are several reasons for this. First, the MAR assumption is general enough that it allows imputation procedures to adjust for differences between the incomplete cases and the complete cases on variables that are observed. Second, inclusion of a large number of predictors in the imputation model, as was done in this project (see Section 4.2.2), makes the MAR assumption more plausible [12, 36–39]. Finally, because the missing values are not observed, there is no direct information for assessing the MNAR assumption, and procedures based on this assumption can be sensitive to the form of the model used [12, 36, 40].

As discussed in Section 2.2, a major known reason for missingness of DXA values is truncal adiposity 'noise', for which the imputation model accounts to a large extent via the inclusion of variables such as waist circumference, the ratio of waist circumference to height, and BMI as predictors. Missingness is also strongly related to age (due in part to the higher prevalence of non-removable objects in scans for older people), for which the imputation model also adjusts. Moreover, the important predictors of DXA missingness just mentioned, as well as other strong non-DXA predictors of the DXA values and their missingness, have low rates of missingness themselves (see Table IV).

5. Evaluating the imputation procedure

As is the case with any adjustment for missing data, the validity of results with imputation depends on the validity of the model used. Moreover, it is desirable for the imputations to have face validity, because the NHANES data are scrutinized by a large variety of users and because face validity of the imputations is consistent with the validity of the imputation model. Therefore, evaluating the model and imputed data is an important step in the imputation process. Of course, model checking is particularly difficult in the case of missing data, because the data being modeled are not observed. Nevertheless, three useful types of diagnostics for imputations are suggested by Abayomi *et al.* [41], namely checks of the fit of the imputation model to the observed data; displays of the completed data to check for unusual patterns; and comparisons of the distributions of observed and imputed data values. Section 5.1 describes diagnostics of the first type, and Section 5.2 describes those of the second and third types.

5.1. Evaluating the regression models

Plots of the residuals from fitting the regression models used in the SRMI procedure to the complete cases generally indicated good fits, little evidence of heteroscedasticity, and nearly normal error distributions. However, plots of the residuals versus the fitted values for some of the head measurement variables displayed 'fan shapes', indicating possible heteroscedasticity. The estimated Box–Cox [13] transformations for these variables tended to be quite strong as well. Re-fitting the regression models with untransformed or logarithm-transformed head variables did not improve the fanning in the plots; hence, the estimated Box–Cox transformations were retained. Although heteroscedasticity could result in some implausibly low imputed values, the lower bounds placed on the imputations would counteract this effect.

Influential observations in the regression models were identified from plots of the residuals and from the review of the numerical diagnostic DFFITS [42]. Sensitivity of the imputations to removal of the influential observations was assessed by re-implementing the imputation procedure with the influential observations removed. If lower bounds were not imposed on the imputations, removal of the influential observations occasionally decreased the prevalence of implausibly low values for some variables but increased the prevalence for other variables. Removal of the influential observations also occasionally lowered what were considered to be reasonably high imputed values for still other variables. These phenomena are not surprising, given the effects that removal of influential observations can have on fitted regressions, and given the highly multivariate nature of this problem. For most variables, however, removal of the influential observations made little difference. Therefore, the influential observations were retained in the imputation procedure, with the setting of lower bounds resolving the issue of implausibly low imputed values.

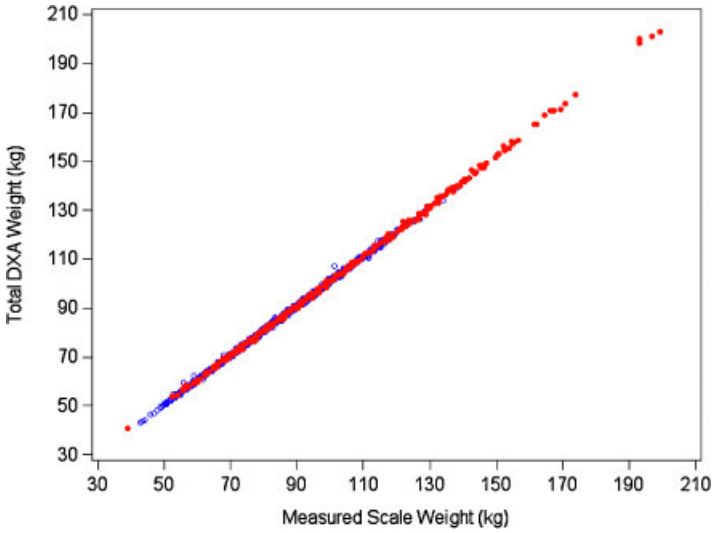

Figure 1. Total weight from dual-energy x-ray absorptiometry (DXA) versus measured scale weight, for completed data based on the first set of imputations, males of ages 20–39, National Health and Nutrition Examination Survey, 1999–2004. Key: blue, measured; red, components imputed.

5.2. Evaluating the imputed values

To assess whether the imputed DXA values might have disproportionately high survey weights and thus be overly influential, the distributions of the survey weights for the non-missing and imputed values were compared. No systematic relations were observed between the survey weights and missingness.

A later analysis of the relations between clustering and missingness was performed as well. A fitted logistic regression with a response variable for whether a person's DXA data are incomplete and explanatory indicator variables for the PSUs showed a highly statistically significant relation. However, when such logistic regressions were fitted within the age groups used for imputation, and other variables from the imputation model were included (e.g. BMI, ratio of waist circumference to height, race/ethnicity, income, education, and survey weight), the statistical significance of the PSUs either decreased substantially or disappeared. This was the case although variables for region, metropolitan-area status, and data release cycle, which were included in the imputation model to reflect the PSUs, could not be included in the logistic regressions due to their perfect collinearity with the PSU indicators. Perhaps most important, however, the missingness rates for the 87 PSUs are between 10 and 36 per cent, with only five of them larger than 30 per cent. Thus, none of the PSUs is overly influential in the sense of having a majority of its values imputed.

The imputed values themselves were examined in several different ways to assess whether they appeared reasonable. For example, total DXA weight, calculated as the sum over the DXA regions, was highly correlated with measured body weight from the NHANES physical examination for each gender-age group, suggesting that the imputations were accurate and precise; see, for example, Figure 1.

Plots of DXA total fat or trunk fat versus variables, such as measured BMI, body weight, waist circumference, and waist circumference divided by height showed that the relations involving the imputed DXA values were similar to those involving the non-missing DXA values; see, for example, Figure 2. (The measured variables against which the DXA fat variables were plotted are highly correlated; hence, the plots discussed here tended to look similar across the measured variables.) The plots also reflected the fact that imputed values for participants with body weight greater than 300 pounds (who were not scanned) and imputed trunk fat values for participants with missing data due to truncal adiposity 'noise' were extrapolated beyond the observed values, as would be expected. Plots of DXA values for left limbs and ribs versus right limbs and ribs also showed that the relations involving the imputed data were similar to those involving the non-missing data; see, for example, Figure 3.

As a final example, comparisons of numerical summaries of imputed and non-missing DXA values tended to produce plausible results, especially in light of the fact that participants with missing DXA values were known to often have different characteristics from participants with non-missing values. For example, large maximum imputed values were found to belong to participants with high observed BMIs and were judged to be reasonable. Similarly, a few very small imputed values seen across all five data files were found to belong to participants with low observed BMIs and also were judged to be reasonable.

Published in 2010 by John Wiley & Sons, Ltd.
Statist. Med. ***2011***, 30 260–276

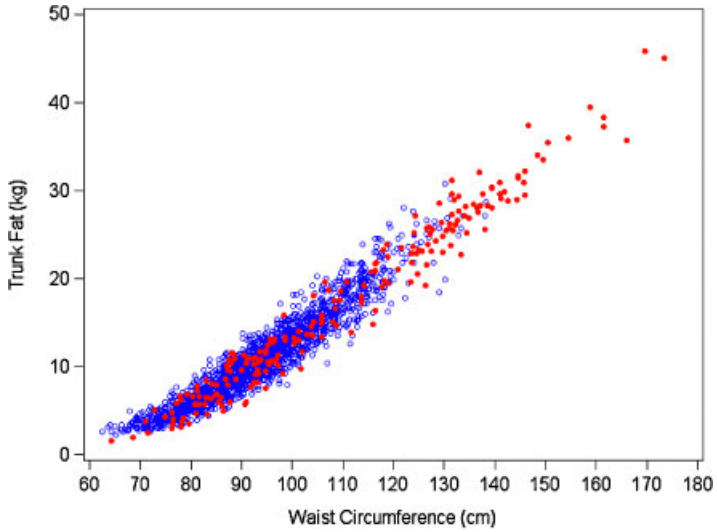

Figure 2. Trunk fat from dual-energy x-ray absorptiometry versus waist circumference, for completed data based on the first set of imputations, males of ages 20–39, National Health and Nutrition Examination Survey, 1999–2004. Key: blue, measured; red, imputed.

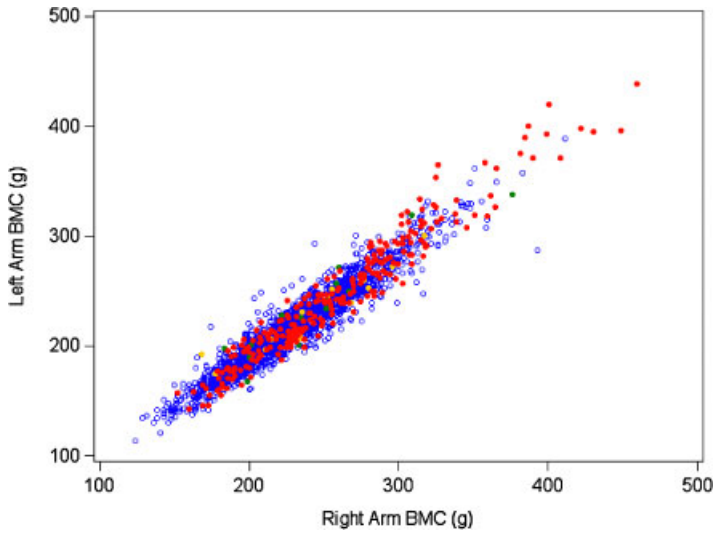

Figure 3. Bone mineral content for the left arm versus bone mineral content for the right arm, for completed data based on the first set of imputations, males of ages 20–39, National Health and Nutrition Examination Survey, 1999–2004. Key: blue, measured for both arms; yellow, measured for left arm and imputed for right arm; green, measured for right arm and imputed for left arm; red, imputed for both arms.

6. Examples of analyses of the DXA data

To illustrate the properties of the multiply imputed DXA data, particularly those properties foreshadowed in Section 1, examples of analyses of the data are presented. SUDAAN software [43] was used to compute point estimates and estimated standard errors, accounting for the complex sample design of the NHANES. The SUDAAN computations for multiply imputed data followed the standard procedures described in [12, 44].

Table V presents estimates of mean body fat percentage for adults of ages 40–59, by gender and BMI category. Table VI presents estimated percentages of the population above a cut point for body fat percentage for the same groups as in Table V. (Other recent analyses of data on body fat from DXA and BMI are reported by Flegal *et al.* [45].) The analysis problems considered in Table VI mimic a frequent type of analysis, in which the percentage of a population or subpopulation above a standard cut point for a variable is estimated. Because there are no standard cut points for body fat percentage, however, arbitrary cut points, 30 per cent for males and 40 per cent for females, were chosen to be near the

Published in 2010 by John Wiley & Sons, Ltd.

Statist. Med. ***2011***, 30 260–276

Statistics in Medicine

Table V. Estimated means of body fat percentage, by gender and body mass index*, for persons of ages 40–59 years, without imputation and with multiple and single imputation, National Health and Nutrition Examination Survey, 1999–2004.

			Without imputation				With imputation					
									Multiple		Single	
	Gender (1)	Body mass index (kg/m^2) (2)	Sample size (3)	Percentage of sample (4)	Estimated mean (5)	Estimated standard error (6)	Sample size (7)	Percentage of Sample (8)	Estimated mean (9)	Estimated standard error (10)	Estimated mean (11)	Estimated standard error (12)
(1)	Male	Total	1,599	100.0	28.0	0.14	2011	100.0	28.7	0.16	28.7	0.16
(2)		<20	50	3.1	20.0	0.55	60	3.0	19.9	0.53	19.8	0.54
(3)		20–<25	383	24.0	24.1	0.27	449	22.3	23.9	0.25	24.0	0.25
(4)		25–<30	762	47.7	28.0	0.18	869	43.2	28.0	0.17	28.0	0.17
(5)		30–<35	325	20.3	31.7	0.21	417	20.7	31.8	0.18	31.8	0.19
(6)		35–<40	74	4.6	35.7	0.39	141	7.0	35.7	0.30	35.7	0.27
(7)		40+	5	0.3	36.5	0.70	75	3.7	38.9	0.50	38.8	0.43
(8)	Female	Total	1584	100.0	39.7	0.23	2035	100.0	40.5	0.22	40.5	0.22
(9)		<20	66	4.2	29.2	0.42	83	4.1	29.2	0.52	29.0	0.47
(10)		20–<25	421	26.6	35.3	0.29	499	24.5	35.4	0.26	35.3	0.25
(11)		25–<30	525	33.1	40.7	0.20	609	29.9	40.8	0.17	40.8	0.17
(12)		30–<35	364	23.0	43.9	0.19	430	21.1	44.0	0.19	44.0	0.18
(13)		35–<40	168	10.6	46.7	0.30	237	11.6	46.8	0.28	46.8	0.25
(14)		40+	40	2.5	48.8	0.57	177	8.7	50.0	0.31	49.8	0.28

*The calculations for this table did not include persons (2.4 per cent) for whom body mass index is missing.

Published in 2010 by John Wiley & Sons, Ltd.

Table VI. Estimated percentages above a cut point* for body fat percentage, by gender and body mass index†, for persons of ages 40-59 years, without imputation and with multiple and single imputation, National Health and Nutrition Survey, 1999–2004.

			Without imputation				With imputation					
									Multiple		Single	
	Gender (1)	Body mass index (kg/m^2) (2)	Sample size (3)	Percentage of sample (4)	Estimated percentage above cut point (5)	Estimated standard error (6)	Sample size (7)	Percentage of sample (8)	Estimated percentage above cut point (9)	Estimated standard error (10)	Estimated percentage above cut point (11)	Estimated standard error (12)
(1)	Male	Total	1599	100.0	34.4	1.55	2011	100.0	39.5	1.44	39.5	1.39
(2)		<20	50	3.1	0.0	0.00	60	3.0	0.0	0.00	0.0	0.00
(3)		20–<25	383	24.0	7.9	1.64	449	22.3	7.3	1.52	7.5	1.46
(4)		25–<30	762	47.7	29.2	2.32	869	43.2	29.5	2.19	29.6	2.14
(5)		30–<35	325	20.3	66.1	3.23	417	20.7	67.9	2.78	67.6	2.73
(6)		35–<40	74	4.6	97.6	1.34	141	7.0	98.3	1.01	98.5	0.72
(7)		40+	5	0.3	100.0	0.00	75	3.7	100.0	0.00	100.0	0.00
(8)	Female	Total	1584	100.0	52.4	1.78	2035	100.0	56.6	1.52	56.6	1.49
(9)		<20	66	4.2	0.0	0.00	83	4.1	0.6	0.81	1.0	0.96
(10)		20–<25	421	26.6	13.7	2.06	499	24.5	15.1	1.91	15.4	1.80
(11)		25–<30	525	33.1	59.7	3.28	609	29.9	60.4	2.92	60.0	2.92
(12)		30–<35	364	23.0	89.6	1.64	430	21.1	90.0	1.57	89.8	1.57
(13)		35–<40	168	10.6	97.9	1.18	237	11.6	98.5	0.82	98.5	0.77
(14)		40+	40	2.5	100.0	0.00	177	8.7	100.0	0.00	100.0	0.00

*The cut points used for body fat percentage were 30 per cent for males and 40 per cent for females.

†The calculations for this table did not include persons (2.4 per cent) for whom body mass index is missing.

Published in 2010 by John Wiley & Sons, Ltd.

Statistics in Medicine

centers of the overall distributions for males and females. The rows and columns of Tables V and VI are labeled for reference in the text that follows.

Point estimates and estimated standard errors are presented for the data without imputation (columns 5 and 6), with multiple imputation (columns 9 and 10), and with only the first set of imputations (columns 11 and 12). Properties of these estimates hold, with occasional exceptions, for results of analyses for other age groups (not shown here).

6.1. Comparison of imputation to no imputation

Because missingness of DXA data is more frequent at higher BMI levels than at lower BMI levels, the prevalence of higher BMI levels is lower among participants with no missing DXA values than it is for the general population. Thus, the estimates based on the DXA data without imputation tend to be biased toward values corresponding to lower BMI levels.

The smaller percentages of observations in the unimputed data at higher BMI levels can be seen in either Tables V or VI. For example, the percentages of men that fall into the 35–<40 and 40+ BMI categories are 4.6 and 0.3, respectively, for men with no imputation (rows 6 and 7, column 4), versus 7.0 and 3.7, respectively, for all men (rows 6 and 7, column 8). Correspondingly, the estimates of mean body fat percentage (Table V) and the percentage of the population above a cut point for body fat percentage (Table VI), without conditioning on BMI level, tend to be lower without imputation (rows 1 and 8, column 5) than with imputation (rows 1 and 8, column 9 or 11). The differences between estimates with imputation and those without imputation are larger in Table VI than in Table V, because the estimated percentage above a cut point is influenced more by the high imputed body fat values for those with high BMI levels than is the estimated mean. This illustrates that the effect of imputation varies depending on the analysis of interest.

The differences in Tables V and VI between estimates based on the data with imputation (columns 9 and 11) and estimates based on the unimputed data (column 5) also tend to be smaller when the analysis is conditioned on BMI category (rows 2–7, 9–14) than when it is not (rows 1 and 8), because conditioning on BMI category removes much of the effect of lower BMI levels in the unimputed data. Two exceptions are the estimates in Table V for males and females with BMI values of 40 or greater (rows 7 and 14). Because this BMI category has especially high levels of missing data and is also unbounded above, there remains a tendency for the data to have higher BMI values with imputation than without imputation.

It can be misleading to compare estimated standard errors of two point estimates when one point estimate is biased, because a biased point estimate can imply a biased estimate of standard error (e.g. the standard error of a proportion is a function of the value of the proportion). However, some tentative conclusions can be drawn from Tables V and VI about the differences between the estimated standard errors with multiple imputation (column 10) and without imputation (column 6) by conditioning on BMI category (rows 2–7, 9–14), so that differences in point estimates with and without imputation are decreased. The estimated standard errors with multiple imputation are smaller than those without imputation 18 out of 20 times in Tables V and VI (situations with estimated percentages of 0 or 100 are omitted). The ratios of the former to the latter range from 0.54 to 1.25 and tend to be smaller when there is a large amount of missing data (as reflected by the smaller sample size without imputation relative to that with imputation—see columns 3 and 7, respectively). The tendency toward lower estimated standard errors with imputation is due to the information added by including the participants with imputed values in the analysis. Such information comes from the observed DXA values for those participants as well as the additional predictors used in the imputation process.

6.2. Comparison of multiple imputation to single imputation

The point estimates for both multiple imputation and single imputation in Tables V and VI (columns 9 and 11) are unbiased under the assumption that the imputation model is correct. However, the multiple-imputation estimates are more efficient, because they are averages of completed-data estimates across five imputations. Thus, the true (not estimated) standard errors for single imputation are higher than those for multiple imputation. In contrast, the estimated standard errors for single imputation (column 12), which do not incorporate the uncertainty due to imputation and are therefore biased downward, are smaller than those for multiple imputation (column 10) 19 out of 25 times (again, situations with estimated percentages of 0 or 100 are omitted). The ratios of the former to the latter range from 0.71 to 1.18. The ratio of 1.18, which is the only one larger than 1.01, occurs in Table VI for females with BMI levels less than 20 (row 9). In this situation, both point estimates of percentages above the cut point (0.6 per cent in column 9 and 1.0 per cent in column 11) are very small, and therefore the estimated standard errors are very sensitive to the values of the point estimates.

6.3. Estimated fractions of missing information

For a given analysis problem, the 'fraction of missing information' [44], often denoted by γ, measures the relative difference between the information (inverse of variance) that would have been contained in the following two

Published in 2010 by John Wiley & Sons, Ltd.
Statist. Med. ***2011***, 30 260–276

analyses: the analysis of the data had there been no missingness and the multiple-imputation analysis when there are missing data. The value of γ is typically less than the missing-data rate, because γ accounts for the predictive power of the imputation model and the predictors included in the model. For many analyses of the multiply imputed DXA data, one would expect γ to be less than 20 per cent, because the missing-data rates for the DXA variables range between 13.8 and 17.5 per cent, and the percentage of persons with incomplete DXA data is about 21 per cent. In addition, the imputation model contains many strong predictors of the DXA components.

The fraction of missing information was estimated via the method of Barnard and Rubin [46] for each analysis problem (i.e. row) in Tables V and VI, with the three situations in Table VI (rows 2, 7, and 14) for which the point estimates are 0 or 100 per cent excluded. Although γ is not estimated reliably when the number of imputations (M) is not large [47, 48], the results were consistent with the expectation expressed in the preceding paragraph. Nineteen out of 25 estimated fractions of missing information, $\hat{\gamma}$, were less than 10 per cent, and three were between 10 and 20 per cent. There were only three values of $\hat{\gamma}$ larger than 20 per cent, namely 34, 40, and 42 per cent. These large values corresponded to situations (Table V, row 7; Table VI, rows 6 and 9) with very high missing-data rates and/or point estimates very close to 0 or 100 per cent, the latter of which would result in volatile estimates of variances that are key components in estimating γ.

7. Discussion

The analyses of body fat presented in Section 6 demonstrated several properties of the multiply imputed DXA data and were consistent with the general assertions made in Section 1. Examples using bone mineral density were also examined (not shown here for brevity). The properties demonstrated by the analyses of bone mineral density were similar in many respects to those demonstrated by the analyses of body fat, especially with regard to comparisons of estimated standard errors. The effects of imputation on point estimates for the analyses of bone mineral density were not in one direction as consistently as for the analyses of body fat, however. This may be due, at least in part, to the countervailing relationships between bone mineral density and factors positively associated with missingness, such as BMI and age (e.g. bone mineral density is positively associated with BMI, but is negatively associated with age).

A common question is whether it is valid to examine the relationship between a variable having some imputed values and another variable that was used as a predictor in the imputation model, as was done in the analytic examples in Section 6. The answer is 'yes', if the imputations were created as random draws from a predictive distribution rather than as deterministic predictions, and if multiple imputation was used to reflect uncertainty about the relationship in the imputed values [30, 31]. In fact, omitting a variable from the imputation model implies that it is known with certainty that the omitted variable has no relationship with the variable being imputed, given the other variables in the model. Unless there is actually no relationship, biases can occur in point estimates based on the data with imputation, such as attenuated estimated relationships between variables [30, 31, 49]. Analyses of data with imputation that involve variables not used in the imputation model can yield biased estimates of variance as well [29, 30, 36].

As expressed in Section 4.2, it is typically advised that it is beneficial to include a large number of predictors in models for multiple imputation [29–32]. One reason for the advice is the potential for the types of bias mentioned in the preceding paragraph. In this regard, the advice is especially relevant for the creation of public-use data, since many different types of analyses will be carried out with such data. The potential decreases in bias due to including a large number of predictors in an imputation model are usually felt to be adequate compensation for potential losses in efficiency due to including possibly irrelevant variables [30].

As can be seen from Figures 1 and 2, some of the imputations for DXA values involved extrapolation beyond the range of the observed data. This was expected and was in fact necessary, given that there were several individuals with missing DXA data who had higher values of variables, such as BMI and waist circumference, than occurred for any of the individuals with observed DXA data. However, the need for such extrapolation also implies that some of the imputations relied on features of the model that were not fully verifiable based on the observed data, although the imputed values and the analyses of them appeared reasonable. Interestingly, some popular types of 'less parametric' imputation procedures, such as hot-deck imputation, which draws the imputed value of a variable for an individual from observed values for other 'donor' individuals with similar values of predictor variables, would not have worked well in the DXA project, because donors would not have been available at the highest levels of predictors such as BMI and waist circumference.

The focus of the NHANES DXA imputation project was on handling missing DXA data. Had missingness of non-DXA variables been a focus as well, a possible alternative approach to the project would have been to use two-stage multiple imputation [50–54], which creates multiple imputations for one type of missing data (say, missingness on the DXA variables) conditionally on each of multiple imputations for the other type of missing data (say, missingness on the non-DXA variables), resulting in a nested structure of multiple imputations. The analysis of two-stage multiply

imputed data is a generalization of the standard multiple-imputation analysis and is not currently programmed into most statistical packages; but it allows an assessment of the separate contributions to uncertainty of the two types of missing data involved (in the current case, the missing data for the non-DXA and DXA variables). Two-stage multiple imputation also allows the possibility of making different assumptions about the missingness of the two types of data and using different procedures to handle them.

Recent work has suggested that the number of imputations of the missing values, M, should generally be larger than the traditional number of five or so. For example, Graham *et al.* [47] found that for statistical testing for small effects, the power falloff for small M relative to $M=100$ is greater than the falloff in relative efficiency traditionally examined (e.g. [55]). In a simulation study in their paper, for testing a small regression coefficient (with a null hypothesis of 0), the respective powers for $M=100$ and $M=5$, by fraction of missing information (γ) were as follows: 0.79 and 0.78 when $\gamma=10$ per cent; 0.79 and 0.73 when $\gamma=30$ per cent; 0.78 and 0.68 when $\gamma=50$ per cent; 0.78 and 0.61 when $\gamma=70$ per cent; and 0.78 and 0.53 when $\gamma=90$ per cent. They argued that although the power falloff is relatively modest when $\gamma\leqslant 30$ per cent, γ itself is not reliably estimated unless M is rather large (see also [48]). Thus, they recommended use of a larger M.

Section 6.3 suggested that the fractions of missing information for the multiply imputed DXA data tend to be moderate (e.g. less than 20 per cent). Consistent with those results, experiences of the authors of this paper to date have indicated that the differences among the analyses of the data with the five different sets of imputations have been small. Given the evidence for low fractions of missing information, it is suspected that the value $M=5$ should generally suffice for the DXA data. Although it would have been more conservative to generate a larger number of imputations, and use of a larger M might well be prudent in other applications, the use of a modest value ($M=5$) has an important potential advantage. The additional computational burden for analyzing multiply imputed data is generally small when there are automated procedures for managing the multiply imputed data sets, conducting the analysis of each data set, and combining the analyses; but such automated procedures are not always available to analysts of the multiply imputed DXA data. Moreover, many analysts typically are not familiar with multiple imputation and could find it to be a nuisance or intimidating. Hopefully, the modest value of M in this application will encourage analysts of the DXA data to actually use all of the imputations, rather than simply picking one and conducting a single-imputation analysis.

This paper described an application of multiple imputation to a major national health survey and thereby provided examples of the types of considerations that arise in such an application. Examples of analyses of the multiply imputed data illustrated important general points, namely, that imputation can help to correct for bias when the missing data are not MCAR; that imputation facilitates the incorporation of information available from the incomplete cases into analyses; and that multiple imputation helps to assess the uncertainty due to imputation. It is hoped that the paper will serve a helpful case study for those considering the use of multiple imputation in other applications as well as a resource for analysts of the multiply imputed DXA data.

Acknowledgements

The authors thank the Associate Editor and two referees for thoughtful comments that helped to improve the manuscript substantially.

References

1. Njeh CF, Fuerst T, Hans D, Blake GM, Genant HK. Radiation exposure in bone mineral density assessment. *Applied Radiation and Isotopes* 1990; **50**:215–236.
2. Wahner HW, Fogelman I. *The Evaluation of Osteoporosis*: *Dual Energy X-ray Absorptiometry in Clinical Practice*. Martin Dunitz: London, 1994.
3. World Health Organization. *Assessment of Fracture Risk and its Application to Screening for Post Menopausal Osteoporosis*. WHO Technical Report Series, vol. 843. World Health Organization: Geneva, 1994.
4. Genant HK, Engelke K, Fuerst T, Glüer CC, Grampp S, Harris ST, Jergas M, Lang T, Lu Y, Majumdar S, Mathur A, Takada M. Noninvasive assessment of bone mineral and structure: state of the art. *Journal of Bone and Mineral Research* 1996; **11**:707–730.
5. Lohman TG, Chen Z. Dual-energy x-ray absorptiometry. In *Human Body Composition* (2nd edn), Heymsfield SB, Lohman T, Wang Z-M, Going SB (eds). Human Kinetics: Champaign, 2005; 63–77.
6. Thomas SR, Kalkwarf HJ, Buckley DD, Heubi JE. Effective dose of dual-energy x-ray absorptiometry scans in children as a function of age. *Journal of Clinical Densitometry* 2005; **8**:415–422.
7. World Health Organization. *Physical Status*: *The Use and Interpretation of Anthropometry*. Report of a WHO Expert Committee. WHO Technical Report Series, vol. 854. World Health Organization: Geneva, 1995.
8. Rubin DB. Inference and missing data (with Discussion). *Biometrika* 1976; **63**:581–592.
9. Little RJA, Rubin DB. *Statistical Analysis with Missing Data* (2nd edn), Section 3.2. Wiley: Hoboken, 2002.
10. Rubin DB. Multiple imputations in sample surveys—a phenomenological Bayesian approach to nonresponse. *Proceedings of the Section on Survey Research Methods*. American Statistical Association: Alexandria, 1978; 20–34.

11. Rubin DB. *Multiple Imputation for Nonresponse in Surveys*. Wiley: New York, 1987.
12. Harel O, Zhou X-H. Multiple imputation: review of theory, implementation and software. *Statistics in Medicine* 2007; **26**:3057–3077. DOI: 10.1002/sim.2787.
13. Box GEP, Cox DR. An analysis of transformations (with Discussion). *Journal of the Royal Statistical Society Series B* 1964; **26**:211–252.
14. National Center for Health Statistics. Analytic and reporting guidelines: the National Health and Nutrition Examination Survey (NHANES). Division of Health and Nutrition Examination Surveys, National Center for Health Statistics, Centers for Disease Control and Prevention, U.S. Department of Health and Human Services, Hyattsville, 2006. Available from: http://www.cdc.gov/nchs/data/nhanes/nhanes_03_04/nhanes_analytic_guidelines_dec_2005.pdf [27 May 2010].
15. Mohadjer L, Curtin LR. Balancing sample design goals for the National Health and Nutrition Examination Survey. *Survey Methodology* 2008; **34**:119–126.
16. Park I, Dohrmann S, Montaquila J, Mohadjer L, Curtin LR. Reducing the risk of data disclosure through area masking: limiting biases in variance estimation. *Proceedings of the Section on Physical and Engineering Sciences*. American Statistical Association: Alexandria, 2006; 1761–1767.
17. National Center for Health Statistics. National Health and Nutrition Examination Survey body composition procedures manual. Division of Health and Nutrition Examination Surveys, National Center for Health Statistics, Centers for Disease Control and Prevention, U.S. Department of Health and Human Services, Hyattsville, 2004. Available from: http://www.cdc.gov/nchs/data/nhanes/nhanes_03_04/BC.pdf [27 May 2010].
18. 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.
19. Kennickell AB. Multiple imputation in the survey of consumer finances. *Proceedings of the Business and Economic Statistics Section*. American Statistical Association: Alexandria, 1998; 11–20.
20. National Center for Health Statistics. Third National Health and Nutrition Examination Survey (NHANES III, 1988–1994): multiply imputed data set, Series 11, No. 7A. Division of Health and Nutrition Examination Surveys, National Center for Health Statistics, Centers for Disease Control and Prevention, U.S. Department of Health and Human Services, Hyattsville, 2001. Available from: http://www.cdc.gov/nchs/nhanes/nh3data.htm [27 May 2010].
21. Schenker N, Raghunathan TE, Chiu P-L, Makuc DM, Zhang G, Cohen AJ. Multiple imputation of missing income data in the National Health Interview Survey. *Journal of the American Statistical Association* 2006; **101**:924–933.
22. Pedlow S, Luke JV, Blumberg SJ. Multiple imputation of missing household poverty level values from the National Survey of Children with Special Health Care Needs, 2001, and the National Survey of Children's Health, 2003. Survey Planning and Special Surveys Branch, Division of Health Interview Statistics, National Center for Health Statistics, Centers for Disease Control and Prevention, U.S. Department of Health and Human Services, Hyattsville, 2007. Available from: http://www.cdc.gov/nchs/data/slaits/mimp01_03.pdf [27 May 2010].
23. Raghunathan TE, Lepkowski JW, Van Hoewyk J, Solenberger P. A multivariate technique for multiply imputing missing values using a sequence of regression models. *Survey Methodology* 2001; **27**:85–95.
24. Raghunathan TE, Solenberger P, Van Hoewyk J. IVEware: Imputation and Variance Estimation Software Users Guide. Survey Research Center, Institute for Social Research, University of Michigan, Ann Arbor, 2002. Available from: http://www.isr.umich.edu/src/smp/ive [27 May 2010].
25. SAS Institute Inc. *Base SAS® 9.1.3 Procedures Guide* (2nd edn). SAS Institute Inc: Cary, 2006.
26. Schafer JL. *Analysis of Incomplete Multivariate Data*, Chapters 3 and 4. Chapman & Hall: London, 1997.
27. van Buuren S, Brand J, Groothuis-Oudshoorn C, Rubin D. Fully conditional specification in multivariate imputation. *Journal of Statistical Computation and Simulation* 2006; **76**:1049–1064.
28. van Buuren S. Multiple imputation of discrete and continuous data by fully conditional specification. *Statistical Methods in Medical Research* 2007; **16**:219–242.
29. Meng X-L. Multiple-imputation inferences with uncongenial sources of input (with Discussion). *Statistical Science* 1994; **9**:538–573.
30. Rubin DB. Multiple imputation after 18+ years (with Discussion). *Journal of the American Statistical Association* 1996; **91**:473–489, 507–517.
31. Little RJA, Raghunathan TE. Should imputation of missing data condition on all observed variables? *Proceedings of the Section on Survey Research Methods*. American Statistical Association: Alexandria, 1997; 617–622.
32. 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.
33. Draper NR, Smith H. *Applied Regression Analysis* (2nd edn). Wiley: New York, 1981; 225–226.
34. National Center for Health Statistics. NHANES 2003–2004 Data Release, January 2006: general information about the NHANES 2003–2004 Laboratory Methodology and public data files. Division of Health and Nutrition Examination Surveys, National Center for Health Statistics, Centers for Disease Control and Prevention, U.S. Department of Health and Human Services, Hyattsville, 2006. Available from: http://www.cdc.gov/nchs/data/nhanes/nhanes_03_04/lab_c_generaldoc.pdf [27 May 2010].
35. Schafer JL. *Analysis of Incomplete Multivariate Data*, Section 2.5. Chapman & Hall: London, 1997.
36. Rubin DB, Schenker N. Imputation and multiple imputation. In *Methods and Applications of Statistics in the Life and Health Sciences*, Balakrishnan N (ed.). Wiley: Hoboken, 2009; 425–440.
37. Schafer JL. *Analysis of Incomplete Multivariate Data*, Section 2.5.2. Chapman & Hall: London, 1997.
38. Collins LM, Schafer JL, Kam CM. A comparison of inclusive and restrictive strategies in modern missing data procedures. *Psychological Methods* 2001; **6**:330–351.
39. Rubin DB, Stern HS, Vehovar V. Handling 'Don't know' survey responses: the case of the Sovenian plebiscite. *Journal of the American Statistical Association* 1995; **90**:822–828.
40. Little RJA, Rubin DB. *Statistical Analysis with Missing Data* (2nd edn), Section 15.1. Wiley: Hoboken, 2002.
41. Abayomi K, Gelman A, Levy M. Diagnostics for multivariate imputations. *Journal of the Royal Statistical Society Series C* 2008; **57**:273–291.
42. Belsley DA, Kuh E, Welsch RE. *Regression Diagnostics*: *Identifying Influential Data and Sources of Collinearity*, Chapter 2. Wiley: New York, 1980.
43. Research Triangle Institute. *SUDAAN Language Manual*, *Release 10.0*. Research Triangle Institute: Research Triangle Park, 2008.

44. Rubin DB. *Multiple Imputation for Nonresponse in Surveys*, Section 3.1. Wiley: New York, 1987.
45. Flegal KM, Shepherd JA, Looker AC, Graubard BI, Borrud LG, Ogden CL, Harris TB, Everhart JE, Schenker N. Comparisons of percentage body fat, body mass index, waist circumference, and waist-stature ratio in adults. *American Journal of Clinical Nutrition* 2009; **89**:500–508.
46. Barnard J, Rubin DB. Small-sample degrees of freedom with multiple imputation. *Biometrika* 1999; **86**:948–955.
47. Graham JW, Olchowski AE, Gilreath TD. How many imputations are really needed? Some practical clarifications of multiple imputation theory. *Prevention Science* 2007; **8**:206–213.
48. Harel O. Inferences on missing information under multiple imputation and two-stage multiple imputation. *Statistical Methodology* 2007; **4**:75–89.
49. 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.
50. Shen ZJ. Nested multiple imputation. *Ph.D. Thesis*, Department of Statistics, Harvard University, Cambridge, MA, 2000.
51. Rubin DB. Nested multiple imputation of NMES via partially incompatible MCMC. *Statistica Neerlandica* 2003; **57**:3–18.
52. Harel O, Schafer JL. Multiple imputation in two stages. *Proceedings of the Federal Committee on Statistical Methodology 2003 Conference*. Office of Management and Budget: Washington, 2003; 91–96. Available from: http://www.fcsm.gov/03papers/Harel.pdf [27 May 2010].
53. Reiter JP, Raghunathan TE. The multiple adaptations of multiple imputation. *Journal of the American Statistical Association* 2007; **102**:1462–1471.
54. Harel O. *Strategies for Data Analysis with Two Types of Missing Values*: *From Theory to Application*. Lambert Academic Publishing: Saarbrücken, Germany, 2009.
55. Rubin DB. *Multiple Imputation for Nonresponse in Surveys*, Section 4.1. Wiley: New York, 1987.

Published in 2010 by John Wiley & Sons, Ltd.
Statist. Med. ***2011***, 30 260–276