STATISTICS IN MEDICINE
Statist. Med. 2004; **23**:3077–3087 (DOI: 10.1002/sim.1911)

Multiple imputation for body mass index: lessons from the Australian Longitudinal Study on Women's Health

Gita D. Mishra[1,*,†] and Annette J. Dobson[2]

[1]*MRC - Human Nutrition Research, Cambridge, U.K.*
[2]*School of Population Health, University of Queensland, Australia*

SUMMARY

In large epidemiological studies missing data can be a problem, especially if information is sought on a sensitive topic or when a composite measure is calculated from several variables each affected by missing values. Multiple imputation is the method of choice for 'filling in' missing data based on associations among variables. Using an example about body mass index from the Australian Longitudinal Study on Women's Health, we identify a subset of variables that are particularly useful for imputing values for the target variables. Then we illustrate two uses of multiple imputation. The first is to examine and correct for bias when data are not missing completely at random. The second is to impute missing values for an important covariate; in this case omission from the imputation process of variables to be used in the analysis may introduce bias. We conclude with several recommendations for handling issues of missing data. Copyright © 2004 John Wiley & Sons, Ltd.

KEY WORDS: multiple imputation; bias correction; body mass index; Australian Longitudinal Study on Women's Health

1. INTRODUCTION

Missing data are a problem, especially for public health research relying on routinely collected databases or large-scale epidemiological studies [1]. In these situations it is often not feasible, for reasons of timing or resources, to ensure completeness and plausibility of every data item collected [2, 3]. Data are especially likely to be missing if they are related to sensitive issues such as alcohol consumption [4, 5] or to intrusive examinations [6].

The problem is exacerbated for composite variables calculated from several items. For example, for the baseline survey (Survey 1) for the mid-aged cohort in the Australian Longitudinal Study on Women's Health (ALSWH) [7] 0–3 per cent of data for individual items from the

*Correspondence to: Gita Mishra, MRC - Human Nutrition Research, Elsie Widdowson Laboratory, Fulbourn Road, Cambridge CB1 9NL, U.K.
†E-mail: gita.mishra@mrc-hnr.cam.ac.uk

Contract/grant sponsor: Australian Commonwealth Department of Health and Ageing

Received February 2003
Accepted May 2004

Medical Outcomes Study SF-36 health-related quality of life measure [8] were missing. The Physical Component Score (PCS) and Mental Component Score (MCS) are calculated from linear combinations of the scores for the individual items after some *ad hoc* imputation rules are applied. For the ALSWH the effect of missing data on individual items is that 6.3 per cent of respondents had missing results for PCS and MCS. This can be a problem both through loss of statistical power (if only complete case data are analysed) and bias (if respondents with missing data differ systematically from respondents providing complete data).

The body mass index (BMI) calculated as weight (in kilograms) divided by the square of height (in meters) is an important variable for predicting health outcomes. For example, incidence of diabetes mellitus, cholecystectomy, and bone and joint disorders increase with BMI, while prevalence of some conditions such as iron deficiency declines and others show U-shaped patterns [9, 10]. The BMI is particularly subject to the missing data problem because it is a composite variable and because some people are especially reluctant to report their weight.

Multiple imputation is the method of choice for handling missing data problems in large and complex data sets [11]. This method involves producing several complete data sets, by randomly sampling likely values for missing items from the estimated multivariate distribution of a group of inter-related variables. Each of the complete data sets is then analysed separately by conventional methods. Finally, the results from the separate analyses are combined to produce estimates with standard errors that correctly reflect the variability in the data.

Imputation methods, including multiple imputation, rely on the assumption that data are missing at random [12]. In practice this may be difficult to determine. Multiple imputation can be used to correct for non-response bias [6]. Schafer [11] has suggested that provided a large number of inter-related variables are used together then multiple imputation can reduce the effects of such bias.

For a large data set that may be accessed by different researchers to investigate a variety of questions there are specific issues for multiple imputation. For example, Faris *et al*. [13] have argued that the complexity of multiple imputation is a disadvantage that may outweigh the gains in statistical power and bias reduction. Additionally it is preferable that the variables to be used in any analysis should be included in the model for multiple imputation but, as discussed by Meng [14] this may not happen, especially when the imputer and analyser are different people.

In this paper, we explore how multiple imputation for a key variable may be simplified by reducing the number of related variables used in the imputation process while still achieving bias reduction in various forms of analysis. We use the BMI variable from Survey 1 for mid-aged women in the ALSWH because: it is a composite variable particularly affected by missing data; the data are clearly not missing completely at random; there are several closely related variables with less missing data; and it is an important predictor of health outcomes whose prevalence in subpopulations is of public health importance.

2. ALSWH

The ALSWH involves three cohorts defined by age at the baseline survey (Survey 1) which was conducted in 1996: young women aged 18–23 years, $n = 14\,762$ respondents; mid-aged women, 45–50 years, $n = 14\,099$; older women, 70–75 years, $n = 12\,767$. Women in the target age groups were randomly selected from the national health insurance database, Medi-

care, which covers everyone in Australia including recent immigrants and refugees. There was purposive over-sampling of women living in rural and remote areas. Details of the recruitment process and response rates have been reported [7]. The study is designed to track the health of women over a period of at least 20 years. It has a particular focus on use of health services, taking a broad perspective of the women's lives and social context. Since Survey 1 each cohort is being re-surveyed every 3 years. The main survey instruments are mailed questionnaires for self-completion. Each survey form has 200–300 items; some are standardized scales (e.g. SF-36 [8]), others are lists of symptoms and diagnoses, health-related behaviour, socio-demographic variables, etc. The procedure for data collection includes a protocol for repeated reminders for non-respondents based on the work of Dillman [15]. As the study covers the whole of Australia it is not possible to conduct face-to-face interviews or to obtain independent objective measures (except on health service use which can be obtained by record-linkage to the Medicare database).

Survey 1 included questions about height (with responses in centimeters or feet and inches), weight (in kilograms or stones and pounds). These data were used to calculate self-reported BMI. There was also an item about women's perception of their weight 'how would you describe yourself now?' (SRWTCAT) with response categories: very underweight, underweight, slightly underweight, average, slightly overweight, overweight, very overweight, don't know. This is a plausible proxy for BMI since women presumably took their height into consideration when categorizing their weight.

Other items in Survey 1 that are related to weight include the following: 'how much would you like to weigh?' (LIKE) with six ordinal response categories; eating to reduce stress (five ordinal categories); frequency of eating takeaway food (three ordinal categories); frequency of going on a diet to lose weight during the last year (five ordinal categories); weight lost in the last 6 months (three ordinal categories); weight gained in the last 6 months (three ordinal categories); menopausal status or use of hormone replacement therapy (HRT) (four nominal categories); smoking status (three nominal categories); level of physical activity (four ordinal categories); highest educational qualification completed (seven ordinal categories); area of residence (urban, rural or remote, based on postcode and collapsed into two categories); age, in years (continuous); and the SF-36 component scores PCS and MCS (both continuous).

3. METHODS OF ANALYSIS

For practical purposes, data that are outside plausible ranges may have to be regarded as missing. For this study we considered height outside (130, 190) cm and weight outside (30, 130) kg and BMI outside (15, 55) to be highly unlikely for mid-aged women (and more likely due to reporting error) and set the values to 'missing'.

In the first stage of the analysis we investigated the extent of missing data and indications of whether BMI data were likely to be missing completely at random (i.e. missingness is unrelated to the missing value or any other variable), missing at random (i.e. missingness is random, conditional on the observed data), or non-ignorable. We examined associations among variables because imputation uses information from subjects with non-missing data on the variables of interest, as well as information on related variables for all subjects, in order to 'fill in' the missing data and thereby create complete data sets. We used 16

variables: BMI, SRWTCAT, LIKE, and the other 13 variables listed in the last paragraph of Section 2.

For the imputation we used the natural logarithm of BMI to achieve a more normally distributed variable. Data were back-transformed before further analyses were performed. To illustrate the sensitivity of multiple imputation to the inclusion or exclusion of related variables, first we imputed missing values using only observed BMI and each one of the 15 other variables used in this paper. Then we imputed missing values using observed BMI values and 14 of the 15 other variables. Differences between the results from this simple and rapid 2-step imputation process provide an understanding of the precise effects of each variable and the relative robustness of the imputation.

Next, to examine the effects of bias on the estimated distribution of BMI in the study population we set BMI to 'missing' for all those women whose self-reported weight category was 'very underweight', 'underweight', 'don't know' and used multiple imputation with all of the other 15 variables or only for the two variables most closely associated with BMI.

We used SAS release 8.2 [16] to impute five data sets for each analysis using PROC MI with the method of Markov chain Monte Carlo (MCMC chain = multiple). Summary statistics for each data set were generated using PROC CORR with output option (type = cov) and the results were combined using PROC MIANALYZE. Variables BMI, PCS, MCS, and age in years were treated as continous, with the exception of menopausal status, the remainder as ordinal variables. Menopausal status was represented as dummy variables.

Finally, to illustrate the potential effects on an analysis of various prior decisions about imputation we used BMI as a covariate in an analysis of socio-demographic predictors of the prevalence of diabetes. This was carried out using logistic regression in PROC GENMOD, which generated parameter estimates and the covariance matrix (with the ODS output statement) for each imputed data set. The results were combined using MIANALYZE procedure with options PARMS and COVB. Several analyses were performed based on: subjects with complete data only; multiple imputation for only BMI, SRWTCAT, and LIKE, but not the other variables used in the analysis; and multiple imputation for BMI, SRWTCAT, LIKE, and all the other variables used in the analysis.

4. RESULTS

Comparisons between self-reported weight category (SRWTCAT) and missing data for height, weight and BMI suggest that BMI values were not missing at random (Table I). In some weight categories the percentage of missing data for BMI is greater than that for height and weight combined due to values outside the plausible range for BMI being set to missing. Weight and BMI were more likely to be missing if the woman categorized herself as overweight or very overweight than if she said she was of average weight. Less expectedly, BMI was also more likely to be missing if the woman reported being very underweight or underweight.

The peculiarities of the women claiming to be very underweight or underweight are also apparent in Table II which summarizes the distribution of BMI for women without missing data for BMI. Those women who claimed to be very underweight had higher self-reported BMI than the group reporting average weight. In contrast the steady increase in BMI from the categories of slightly underweight through to very overweight is as one would expect.

Table I. Self-reported weight category and percentages of missing data for height, weight and BMI.

Self-reported weight category	Number of respondents	Height	Weight	BMI
Very underweight	70	7.1	10.0	18.6
Underweight	140	3.6	3.6	7.9
Slightly underweight	551	2.4	2.4	3.8
Average	3829	3.2	2.1	4.7
Slightly overweight	4440	3.0	2.9	5.0
Overweight	3572	3.2	4.0	6.3
Very overweight	1412	2.6	6.4	7.9
Don't know or missing	85	25.9	24.7	28.2

Table II. Distribution of self-reported BMI tabulated by other variables for women without missing data.

	Number	Mean	Standard deviation	Median	Lower quartile	Upper quartile
Self-reported weight						
Very underweight	57	27.1	9.5	28.4	17.1	33.8
Underweight	129	21.9	5.5	19.4	18.3	24.6
Slightly underweight	530	20.6	2.9	19.9	18.8	21.6
Average	3648	22.0	2.3	21.8	20.5	23.0
Slightly overweight	4216	24.8	2.7	24.5	23.1	26.1
Overweight	3346	28.8	3.7	28.3	26.2	30.9
Very overweight	1300	34.6	5.2	34.1	31.0	37.9
Like to weigh						
Over 5 kg more	333	21.8	3.9	21.1	19.0	23.7
1–5 kg more	101	24.8	5.9	24.9	19.9	28.2
Happy as I am	2645	21.9	3.3	21.3	20.1	22.9
1–5 kg less	4316	23.7	2.5	23.5	22.1	25.0
6–10 kg less	2845	26.9	3.3	26.5	24.8	28.5
Over 10 kg less	2719	32.4	5.2	31.6	28.8	35.2
Eat to relieve reduce stress						
None of the time	4480	24.0	4.3	23.1	21.2	25.7
Little of the time	3833	25.6	4.8	24.6	22.4	27.8
Some of the time	3254	26.9	5.2	25.8	23.3	29.8
Most of the time	1121	29.0	6.1	28.2	24.6	32.5
All of the time	437	30.3	6.7	29.9	25.5	34.3
Take-away food						
$<$Once a month	4914	25.3	5.1	24.2	21.8	27.5
Once a month	4051	25.8	5.1	24.7	22.3	28.3
$>$Once a month	4318	26.5	5.5	25.2	22.7	29.3
Dieting behaviour last year						
Never	7522	24.3	4.6	23.3	21.3	26.2
1–4 times	4236	27.4	5.1	26.2	23.8	30.1
5–10 times	342	29.8	5.9	28.9	25.0	33.3
$>$10 times	172	29.7	5.4	29.3	25.5	32.7
Always on diet	1005	28.8	5.9	27.5	24.5	32.2

Table II. *Continued.*

	Number	Mean	Standard deviation	Median	Lower quartile	Upper quartile
In the last 6 months lost						
⩾5 kg on purpose	1929	28.7	5.7	27.4	24.4	31.9
⩾5 kg without wanting to	634	24.7	5.5	23.3	20.8	27.3
Not lost ⩾5 kg	10574	25.4	5.0	24.3	22.0	27.6
In the last 6 months gained						
⩾5 kg on purpose	152	26.8	6.3	26.1	21.8	30.8
⩾5 kg without wanting to	2896	27.8	5.3	26.7	24.1	30.4
Not gained ⩾5 kg	10066	25.3	5.0	24.2	21.8	27.5
Menopausal status						
Pre	5389	25.4	5.0	24.3	22.0	27.6
Peri	3605	25.8	5.3	24.6	22.2	28.2
Post	1071	25.7	5.3	24.5	22.1	28.1
HRT	3113	26.7	5.5	25.6	22.8	29.8
Smoking						
Current	2382	25.3	5.0	24.4	21.8	27.8
Ex	3700	26.1	5.3	24.9	22.4	28.5
Never	6802	25.9	5.3	24.7	22.3	28.4
Physical activity						
Nil	3620	26.6	5.8	25.4	22.5	29.8
Low	4037	26.0	5.1	24.9	22.4	28.5
Moderate	3361	25.5	5.0	24.6	22.1	27.8
Vigorous	2190	24.9	4.6	23.8	21.7	27.0
Level of education						
No formal qual.	2370	27.0	5.8	25.7	22.9	30.2
School/intermediate cert.	4186	26.1	5.2	24.9	22.4	28.5
Higher school cert.	2210	25.5	5.1	24.5	22.1	27.9
Trade/apprenticeship	461	25.5	4.7	24.6	22.3	28.0
Certificate/diploma	2077	25.5	5.1	24.2	22.0	27.7
University degree	1198	24.6	4.6	23.5	21.5	26.6
University higher degree	666	24.9	4.9	23.9	21.4	27.1
Area of residence						
Urban	4872	25.4	5.1	24.2	22.0	27.7
Rural/remote	8379	26.1	5.3	24.9	22.3	28.7
Age, years						
45	674	26.0	5.6	24.6	22.0	28.5
46	2786	25.7	5.3	24.5	22.1	28.3
47	2739	25.7	5.2	24.6	22.2	28.2
48	2549	25.6	5.1	24.5	22.1	28.2
49	2567	26.0	5.2	24.9	22.4	28.5
50	1969	26.2	5.3	25.0	22.5	28.7
PCS quintile						
1	2485	27.5	6.3	26.2	23.0	30.7
2	2474	26.7	5.6	25.5	22.7	29.5
3	2505	25.8	4.9	24.8	22.4	28.3
4	2523	24.9	4.3	24.2	21.9	27.1
5	2509	24.4	4.2	23.7	21.5	26.4

Table II. *Continued.*

	Number	Mean	Standard deviation	Median	Lower quartile	Upper quartile
MCS quintile						
1	2456	26.1	5.4	24.9	22.3	29.0
2	2517	25.9	5.5	24.6	22.2	28.4
3	2502	25.6	4.9	24.6	22.1	28.1
4	2514	25.6	5.0	24.6	22.2	27.9
5	2507	26.0	5.4	24.8	22.3	28.6

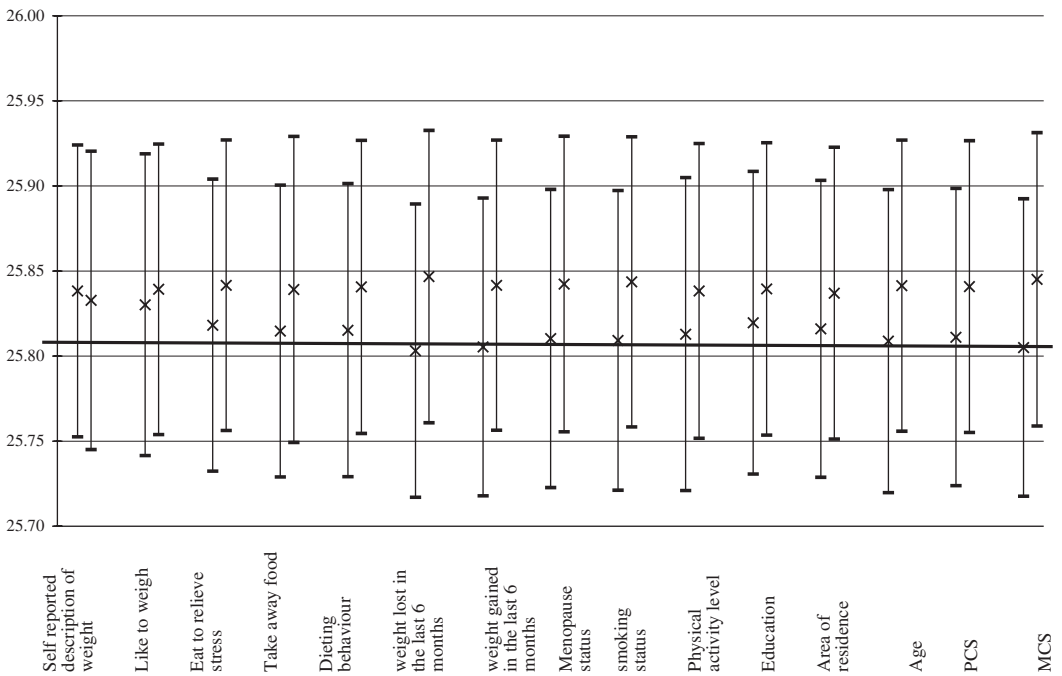

Figure 1. Mean BMI, with 95 per cent confidence interval, obtained by multiple imputation displayed in relation to other variables: the first values are from imputation using only BMI and the other variable, the second values are from imputation using all variables except the one identified. The bold line refers to mean BMI from the data set without missing BMI.

Other variables that appear to be related to BMI, based on results for women with complete data, were: how much they said they would like to weigh (LIKE), eating to reduce stress, frequency of dieting, weight change in the last 6 months, level of physical activity, and educational attainment.

The effects of including or excluding variables in the multiple imputation process are illustrated in Figure 1. The bold horizontal line denotes the mean BMI for those women with complete data. The imputed mean BMI is 25.85 when the whole data set is used for

Table III. Mean (standard deviation) of BMI if all BMI values are regarded as missing for women who said they were very underweight, underweight, gave the response 'don't know' or had missing data for self-reported weight category; using the complete data set only, or data imputed using the three variables BMI, SRWTCAT and LIKE, or data imputed using all 16 weight-related variables.

Self-reported description of weight	Complete data set $n = 13\,226$	Multiple imputation with 3 variables $n = 14\,099$	Multiple imputation with 16 variables $n = 14\,099$
Very underweight	27.13 (9.50)	16.24 (1.68–2.18)	16.87 (1.77–1.98)
Underweight	21.93 (5.51)	17.57 (2.00–2.30)	17.80 (1.96–2.54)
Slightly underweight	20.60 (2.90)	20.56 (2.88–2.89)	20.57 (2.88–2.90)
Average	21.96 (2.29)	21.96 (2.29–2.31)	21.96 (2.30–2.31)
Slightly overweight	24.82 (2.73)	24.85 (2.74–2.75)	24.85 (2.73–2.76)
Overweight	28.82 (3.72)	28.83 (3.69–3.71)	28.84 (3.68–3.71)
Very overweight	34.64 (5.19)	34.51 (5.10–5.15)	34.51 (5.10–5.14)

imputation. For the two variables on the left of Figure 1, SRWTCAT and LIKE, mean BMI values are fairly close whether the variable is the only one used for imputation or the only one excluded. Thus these variables alone provide almost as much information for the imputation of BMI as the whole data set does. In contrast variables at the right end of the figure have little effect: using each with only BMI produces a mean close to that for women without missing data (bold line) or leaving each out produces a mean close to that from the full imputation for all women. Level of education, area of residence and eating to relieve stress all have some effect on imputation of BMI, but less than SRWTCAT and LIKE.

Next we chose to regard BMI as missing for these women in response categories 'very underweight' or 'underweight' for SRWTCAT. Thus we gave more credence to SRWTCAT than to BMI calculated from self-reported weight and height among this group of middle-aged women. We performed multiple imputation first using BMI, SRWTCAT, and LIKE only, and then using all 16 weight related variables. The results are shown in Table III as the mean BMI for each category of SRWTCAT. The effect of multiple imputation is that mean BMI increases across the ordinal categories of SRWTCAT (due to the linearity assumptions implicit in the multivariate normal model used for multiple imputation). Thus imputed BMI means for the 210 women (1.5 per cent of the total) who said they were 'very underweight' or 'underweight' correspond to their perceptions, but differ considerably from BMI means calculated for women in these categories who provided self-reported weights and heights. As expected from Figure 1 using all 16 variables for imputation did not produce results that differed much from those based only on BMI, SRWTCAT, and LIKE.

To examine the effects of using BMI based on multiple imputation as a covariate in another analysis, we used logistic regression to estimate the effects of various risk factors for diabetes mellitus. Table IV shows the prevalence of diabetes and estimated odds ratios for categories of BMI, country of birth and two socio-economic measures for three data sets: women with complete data; multiple imputation data sets with BMI set to missing for women who said they were 'very underweight' or 'underweight' and BMI, SRWTCAT, and LIKE in the imputation model (model A); as for the previous data sets but with *diabetes*, *country of birth*, *ease of managing on their income*, and *occupation* also included in the imputation model (model B).

Table IV. Prevalence of diabetes (as a percentage) and odds ratios and 95 per cent confidence intervals (CI) for diabetes using the complete data set only, or data imputed using BMI, SRWTCAT and LIKE variable (model A), or data imputed using these three variables and all the variables* used in the analysis (model B).

	Complete data set		Imputed model A		Imputed model B	
Variables	per cent	OR (95 per cent CI)	per cent	OR (95 per cent CI)	per cent	OR (95 per cent CI)
BMI						
<20	1.36	1.00 (0.55–1.83)	1.74	0.97 (0.53–1.75)	1.60	0.88 (0.49–1.59)
$\geqslant 20 - \leqslant 25$	1.35	Reference	1.42	Reference	1.35	Reference
$>25 - \leqslant 30$	3.17	2.40 (1.78–3.23)	3.16	2.27 (1.69–3.05)	3.16	2.31 (1.74–3.06)
$>30 - \leqslant 40$	6.02	4.73 (3.51–6.37)	6.02	4.54 (3.39–6.10)	6.17	4.42 (3.33–5.87)
>40	14.12	11.0 (7.04–17.10)	14.18	10.75 (6.91–16.70)	14.34	11.27 (7.41–17.16)
Country of Birth						
Australia	2.77	Reference	2.77	Reference	2.76	Reference
Other English speaking	2.04	0.85 (0.59–1.23)	2.04	0.87 (0.61–1.23)	2.04	0.85 (0.60–1.21)
Europe	3.77	1.25 (0.84–1.86)	3.77	1.17 (0.79–1.75)	3.77	1.28 (0.89–1.84)
Asia	6.84	4.23 (2.70–6.60)	6.84	4.01 (2.60–6.20)	6.78	3.72 (2.43–5.69)
Other	5.88	1.83 (1.02–3.28)	5.88	1.77 (1.01–3.11)	5.67	2.03 (1.24–3.30)
Manage on income						
Impossible/difficult	3.70	Reference	3.70	Reference	3.67	Reference
Not too bad/easy	2.34	0.73 (0.59–0.91)	2.34	0.72 (0.58–0.89)	2.32	0.74 (0.60–0.91)
Occupation						
Manager/professional	2.06	Reference	2.06	Reference	2.07	Reference
Para professional/trade	2.62	1.17 (0.79–1.72)	2.62	1.15 (0.79–1.67)	2.75	1.19 (0.84–1.70)
Admin/sales/service	2.38	1.14 (0.85–1.53)	2.38	1.10 (0.83–1.47)	2.45	1.12 (0.84–1.49)
Machine operator/manual	4.92	2.01 (1.47–2.76)	4.92	1.93 (1.42–2.63)	5.06	1.97 (1.46–2.68)
Other	5.86	1.79 (1.09–2.95)	5.86	2.04 (1.29–3.23)	6.01	2.06 (1.33–3.21)

*Country of birth, manage on income, occupation.

Differences between the results from the complete case data set and model A were mainly for women with low BMI. Confidence intervals from model A were narrower than from the complete data set, reflecting the larger effective sample size. Model A used data imputed from a smaller data set of variables than used in the analysis; model B used data imputed from all the variables in the analysis together with those needed for imputing BMI. The results from the two analyses are different with model A underestimating the effects of BMI in particular. In this example, the most correct analysis (model B) produced results that were quite similar to those from the complete data set.

5. DISCUSSION

Our example illustrates several practical points about multiple imputation. If data are completely missing at random then for a large data set (where lack of statistical power is not a serious concern) multiple imputation may not be worthwhile (see also Reference [1]). However, when data are missing at random (conditional on the observed data) multiple imputation can be a valuable tool for estimating the extent of non-response bias.

For items of a sensitive nature that are likely to be missing it is desirable to collect related items that are more likely to be completed so that these can be used to strengthen the imputation of missing data. Asking women how they classified their weight and other questions related to eating behaviour and weight provided additional information from which we were able to impute BMI. The form of sensitivity analysis illustrated in Figure 1 provides a means of identifying which variables make the greatest contribution to the imputation of a variable of particular interest (in our case, BMI). This graphical method helps the analyst to reduce the number of variables that need to be included in the multiple imputation phase of an analysis and to identify subsets of related variables that should be imputed together. Using partial correlation coefficients or principal components analysis from the complete case data set may be a useful first step. However when a large proportion of the data is missing the selection process may be less than clear-cut. By working with the imputed data sets, the graphical method allows the analyst to see directly the impact of a particular variable on the imputation.

Nevertheless our example about risk factors for diabetes illustrates the need to include in the imputation all those variables that will be used in the analysis. For large data and complex data sets it may be tempting to separate the imputation and analysis phases. But there are dangers if the imputation phase relies only on closely inter-related variables, which will provide most information about missing data, but omits variables that are important for analysis [14].

We conclude with several recommendations:

1. It is worthwhile to use several measures of the same underlying concept or condition. This purposeful redundancy is a basic principle of psychometrics that is often ignored in epidemiological and clinical studies. Multiple measures are valuable for imputation of missing values, particularly for sensitive issues where bias and missingness can be anticipated.
2. The methods section of a substantive paper should describe the imputation procedure used.

3. For a large and complex data set it is likely to be necessary to identify subsets of closely related variables to be used for imputation [1]. Methods such as principal components analysis provide a starting point. Sensitivity analysis using multiple imputation (as in Figure 1) can be employed to refine the subsets of imputation variables.
4. It is advisable to use all the variables required for an analysis in the multiple imputation process. But additional imputation variables, closely related to one or more of the analysis variables, may be beneficial for improving the validity of the results.

ACKNOWLEDGEMENTS

The Australian Longitudinal Study on Women's Health, which was conceived and developed by groups of inter-disciplinary researchers at the Universities of Newcastle and Queensland, is funded by the Australian Commonwealth Department of Health and Ageing. We thank all the participants for their valuable contribution to the project.

REFERENCES

1. Arnold AM, Kronmal RA. Multiple imputation of baseline data in the Cardiovascular Health Study. *American Journal of Epidemiology* 2003; **157**:74–84.
2. Zhou XH, Eckert GJ, Tierney WM. Multiple imputation in public health research. *Statistics in Medicine* 2001; **20**:1541–1549.
3. Perez A, Dennis RJ, Gil JFA, Rondon MA, Lopez A. Use of the mean, hot deck and multiple imputation techniques to predict outcome in intensive care unit patients in Columbia. *Statistics in Medicine* 2002; **21**: 3885–3896.
4. Longford NT, Ely M, Hardy R, Wadsworth MEJ. Handling missing data in diaries of alcohol consumption. *Journal of the Royal Statistical Society Series A* 2000; **163**:381–402.
5. Gmel G. Imputation of missing values in the case of a multiple instrument measuring alcohol consumption. *Statistics in Medicine* 2001; **20**:2369–2381.
6. Taylor JMG, Cooper KL, Wei JT, Sarma AV, Raghunathan TE, Heeringa SG. Use of multiple imputation to correct for nonresponse bias in a survey of urologic symptoms among African-American men. *American Journal of Epidemiology* 2002; **156**:774–782.
7. Brown WJ, Bryson L, Byles JE, Dobson AJ, Lee C, Mishra G, Schofield M. Women's Health Australia: recruitment for a national longitudinal cohort study. *Women and Health* 1998; **28**:23–40.
8. Ware JE, Snow KK, Kosinski M, Gandek B. *SF-36 Health Survey Manual and Interpretation Guide*. The Health Institute, New England Medical Center: Boston, MA, 1993.
9. Brown WJ, Dobson AJ, Mishra G. What is a healthy weight for middle-aged women. *International Journal of Obesity* 1998; **22**:520–528.
10. Brown WJ, Mishra G, Kenardy J, Dobson A. Relationship between body mass index and well-being in young Australian women. *International Journal of Obesity* 2002; **24**:1360–1368.
11. Schafer JL. *Analysis of Incomplete Multivariate Data*. Chapman & Hall: London, 1997.
12. Rubin DR. *Multiple Imputation for Non-response in Surveys*. Wiley: New York, 1987.
13. Faris PD, Ghali WA, Brant R, Norris CM, Galbraith PD, Knudtson ML, for the APPROACH Investigators. Multiple imputation versus data enhancement for dealing with missing data in observational health outcome analyses. *Journal of Clinical Epidemiology* 2002; **55**:184–191.
14. Meng XL. Multiple-imputation inferences with uncongenial sources of input. *Statistical Science* 1994; **9**: 538–573.
15. Dillman DA. *Mail and Telephone Surveys*: *The Total Design Method*. Wiley: New York, 1978.
16. SAS Institute Incorporated. *SAS 8*. SAS Institute Inc.: Cary, NC, 1999.

