American Journal of Epidemiology

Vol. 172, No. 4
DOI: 10.1093/aje/kwq137
Advance Access publication: July 8, 2010

## Practice of Epidemiology

# Strategies for Multiple Imputation in Longitudinal Studies

**Michael Spratt, James Carpenter, Jonathan A. C. Sterne, John B. Carlin, Jon Heron, John Henderson, and Kate Tilling***

* Correspondence to Dr. Kate Tilling, Department of Social Medicine, Faculty of Medicine and Dentistry, University of Bristol, Canynge Hall, 39 Whatley Road, Bristol BS8 2PS, United Kingdom (e-mail: kate.tilling@bristol.ac.uk).

*Initially submitted April 8, 2009; accepted for publication April 20, 2010.*

Multiple imputation is increasingly recommended in epidemiology to adjust for the bias and loss of information that may occur in analyses restricted to study participants with complete data (''complete-case analyses''). However, little guidance is available on applying the method, including which variables to include in the imputation model and the number of imputations needed. Here, the authors used multiple imputation to analyze the prevalence of wheeze among 81-month-old children in the Avon Longitudinal Study of Parents and Children (Avon, United Kingdom; 1991–1999) and the association of wheeze with gender, maternal asthma, and maternal smoking. The authors examined how inclusion of different types of variables in the imputation model affected point estimates and precision, and assessed the impact of number of imputations on Monte Carlo variability. Inclusion of variables associated with the outcome in the imputation model increased odds ratios and reduced standard errors. When only 5 or 10 imputations were used, variability due to the imputation procedure was substantial enough to affect conclusions. Careful preliminary analysis identified the scope for multiple imputation to reduce bias and improve efficiency and provided guidance for building the imputation model. When data are missing, such preliminary analyses should be routinely undertaken and reported, regardless of whether multiple imputation is used in the final analysis.

imputation; longitudinal studies; missing data

Abbreviations: ALSPAC, Avon Longitudinal Study of Parents and Children; FAI, Family Adversity Index; MAR, missing at random; MCAR, missing completely at random; MNAR, missing not at random.

Missing data are ubiquitous in longitudinal studies, because of item, questionnaire, or visit nonresponse and subject attrition (1, 2). It is well established that analyses restricted to study participants with complete data (''complete-case'' analyses), which are the default way of dealing with missing data in statistical software, can be both biased and inefficient. Multiple imputation has been proposed as a remedy for these problems, and its incorporation into routine practice has been recommended (3, 4).

In the presence of missing data, the validity of an analysis depends on the unknown missing-data mechanism and the variables included in the analysis. Missing-data mechanisms are usually characterized using the typology introduced by Rubin (5), as ''missing completely at random'' (MCAR), ''missing at random'' (MAR), or ''missing not at random'' (MNAR). A complete-case analysis is valid if the probability of being a complete case is independent of the outcome variable, given the covariates in the model (6). With standard approaches to multiple imputation, results are valid under the assumption that the data are MAR. However, estimation based on multiple imputation is not always less biased than complete-case analyses (7). For example, complete-case analysis for a model in which a covariate is MNAR but missingness is unrelated to the response will be unbiased. In this situation, an analysis based on multiple imputation can be biased (8, 9).

Multiple imputation allows for uncertainty about the missing data by creating multiple copies of the data set in which missing values are replaced by imputed values sampled from a posterior predictive distribution, itself estimated

from the partially observed data (5, 6, 10–13). Each imputed data set is analyzed and the results are combined, with standard errors that incorporate the variability in results between the imputed data sets (6, 10). Methods for multiple imputation include chained equations (14) and multivariate normal imputation (7) and are implemented in various software packages (15–20).

Results from analyses based on multiple imputation are increasingly being reported in the epidemiologic and medical literature (3). However, little published guidance is available on the choices to be made when using the method, including whether the multiple imputation model should include variables associated with the probability of data being missing, variables predictive of those variables that are subject to missing data, or both. Analysts must also decide how many imputed data sets to create. A small number of imputed data sets has been suggested as adequate (typically 5; for example, see Schafer (21)), but this has recently been questioned (22).

We used multiple imputation to analyze the prevalence of wheeze among 81-month-old children in the Avon Longitudinal Study of Parents and Children (ALSPAC) and associations of wheeze with gender and maternal asthma and smoking. We examined how the choice of imputation model affected estimates of prevalence and association, and how the number of imputations used influenced the Monte Carlo variability of the results. Here we describe the implications for the use of multiple imputation in epidemiology.

## MATERIALS AND METHODS

ALSPAC is a population-based prospective study that included all pregnant women living in Avon, United Kingdom, with an expected date of delivery between April 1, 1991, and December 31, 1992 (23). A total of 14,541 mothers enrolled, and there were 14,062 livebirths; 13,988 of those infants were still alive at 1 year. ALSPAC children have been followed up since recruitment using questionnaires and clinical assessments. The response rate for the 81-month questionnaire was 61% ($n = 8{,}578$). Details on measures, procedures, sample characteristics, and response rates are available on the study's Web site (www.alspac.bris.ac.uk). Ethical approval for the study was obtained from the ALSPAC Law and Ethics Committee and the local research ethics committees.

Our outcome of interest was asthma in childhood. Wheeze in young children is difficult to ascribe to a single diagnostic entity and may be transient. Therefore, we used as our outcome whether the child was reported to have wheezed during the 12 months prior to age 81 months. This was defined as a positive response to either of 2 questionnaire items: 1) whether the child had wheezed since the previous questionnaire (administered at age 69 months) and a doctor had been consulted and 2) whether there had been any periods of wheezing with whistling in the chest when the child breathed. The prognostic variables considered here were gender, maternal asthma, and maternal smoking during pregnancy. Information on maternal history of asthma was obtained from a questionnaire administered early in pregnancy, and information on maternal smoking was obtained from whether a mother reported being a current smoker on a questionnaire filled out in midpregnancy.

The prevalence of wheeze at age 81 months was estimated, with a 95% confidence interval. Mutually adjusted associations of wheeze at age 81 months with gender, maternal asthma, and maternal smoking were estimated using logistic regression. The outcome variable (wheeze at age 81 months) and 2 of the prognostic variables (maternal smoking and maternal asthma) were subject to missing data. We used multiple imputation to address the potential bias and loss of precision that could result from complete-case analysis. Odds ratios estimated using logistic regression should be asymptotically unbiased (in small samples, estimates are systematically biased away from the null (24)).

We considered additional variables not in the analysis model for inclusion in the imputation model. Previous work on ALSPAC suggested that children from lower socioeconomic groups were more likely to drop out of the study (25). Therefore, we used a group of 10 binary socioeconomic status variables on which information was collected during pregnancy (the Family Adversity Index (FAI) (26)), for each of which a code of 1 indicated adverse circumstances. Binary indicators of wheeze since the previous time point at ages 6, 18, 30, 42, 54, and 69 months were created in a manner analogous to wheeze at 81 months. We used 4 separate logistic regression models to examine which of these variables predicted the occurrence of missing data in the outcome variable and each of the 3 prognostic variables, and we used a further logistic regression model to examine which variables were predictive of the values of the variable with the most missing data (wheeze at age 81 months).

We investigated whether the multiple imputation model should include variables associated with the probability of data being missing, variables predictive of those variables that were subject to missing data, or both by comparing 3 imputation models. The first model included all variables found to be related to the probability of missing data on the outcome or any of the prognostic variables (model 1). The second included all of the predictors of values of wheeze at age 81 months (model 2), and the third included both of the above sets of variables (model 3). Each model also included wheeze at age 81 months and the 3 prognostic variables (gender, maternal asthma, and maternal smoking). For each of these 3 models, we used multiple imputation with chained equations (14). In each chained equation cycle, each missing value in each variable is imputed on the basis of a predictive distribution derived from a regression on all other variables in the imputation model. At the end of 10 cycles, 1 imputed data set is created. We repeated this process to create 200 imputed data sets. The results of analyses on each individual data set were combined using Rubin's rules (6, 10). The estimated fraction of missing information was calculated for each model, as the ratio of the between-imputation variance to the sum of the between- and within-imputation variances (27).

Ideally, any analysis should produce the same results if it is repeated on the same data set. The random variability inherent in the imputation process means, in addition to the variability between results of analyses from different

**Table 1.** Twenty Most Frequent Patterns of Missingness of Data on Variables Related to Wheeze at Age 81 Months ($n = 13,983$), Avon Longitudinal Study of Parents and Children, Avon, United Kingdom, 1991–1999[a]

| Presence/Absence of Data on Maternal Asthma | Presence/Absence of Data on Maternal Smoking | FAI[b] | Presence/Absence of Data on Wheeze at the Following Month: | | | | | | | No. of Children | % of Children |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | 6 | 18 | 30 | 42 | 57 | 69 | 81 | | |
| + | + | + | + | + | + | + | + | + | + | 5,494 | 39.3 |
| + | + | − | + | + | + | + | + | + | + | 609 | 4.4 |
| − | + | − | − | − | − | − | − | − | − | 518 | 3.7 |
| + | + | + | + | + | + | + | + | + | − | 409 | 2.9 |
| + | + | + | − | − | − | − | − | − | − | 359 | 2.6 |
| + | + | + | + | + | + | + | + | − | − | 337 | 2.4 |
| + | + | + | + | + | − | − | − | − | − | 309 | 2.2 |
| − | − | − | − | − | − | − | − | − | − | 305 | 2.2 |
| + | + | + | + | + | + | + | + | − | + | 303 | 2.2 |
| + | + | + | + | + | − | + | + | + | + | 295 | 2.1 |
| + | + | + | + | − | − | − | − | − | − | 280 | 2.0 |
| + | + | + | + | + | + | + | − | − | − | 272 | 1.9 |
| + | + | − | − | − | − | − | − | − | − | 255 | 1.8 |
| + | + | + | + | + | + | − | − | − | − | 196 | 1.4 |
| + | + | + | + | + | + | + | − | + | + | 178 | 1.3 |
| + | + | + | + | + | + | − | + | + | + | 122 | 0.9 |
| + | + | + | + | − | + | + | + | + | + | 117 | 0.8 |
| + | + | + | − | + | + | + | + | + | + | 109 | 0.8 |
| + | + | − | + | − | − | − | − | − | − | 107 | 0.8 |
| + | − | − | − | − | − | − | − | − | − | 95 | 0.7 |

Abbreviation: FAI, Family Adversity Index.

[a] A plus sign indicates that the variable was observed; a minus sign indicates that data on the variable were missing. A total of 3,314 children (23.7%) had a missingness pattern other than one of those shown (there were 457 additional missingness patterns not shown in the table).

[b] A plus sign in the FAI column indicates that all FAI variables were present; a minus sign indicates that data on at least 1 FAI variable were missing.

imputed data sets, that the results of multiple imputation *analyses* will vary when repeated. We investigated the extent of this variability by repeating the whole process, varying the number of imputed data sets used (50 replications each of 5, 10, 20, 40, and 200 data sets). These analyses were restricted to the model including both the variables predictive of missingness and the variables predictive of wheeze at age 81 months (model 3 above).

All analyses were performed using Stata, version 10 (Stata Corporation, College Station, Texas) (28).

## RESULTS

A total of 13,983 children were included in the data set. Table 1 shows the most frequent patterns of missing data: 5,494 children (39.3%) had complete data on all variables (outcome, prognostic variables, 10 FAI variables and 6 wheeze variables), while only 409 children (2.9%) had complete data apart from wheeze at age 81 months. The most common patterns of missing data were caused by attrition ($n = 4,045$). A total of 1,532 children had no measurement of wheeze during follow-up. Although these children were included in the analyses, their inclusion made little difference in the results (confirmed by sensitivity analyses).

Table 2 shows the extent of missingness in the outcome variable and the prognostic and FAI variables and the proportion of children with each characteristic. All children had their gender recorded, while 12,303 (88%) of the mothers reported whether they had asthma and 13,163 (94%) reported whether they had smoked during pregnancy. There were 8,402 children (60%) who had the outcome variable observed, while 8,037 (57%) had complete data on the outcome and all 3 prognostic variables.

To investigate which variables predicted missingness, we estimated associations (mutually adjusted) between the outcome, prognostic, and FAI variables and missing data on wheeze at age 81 months, maternal asthma, and maternal smoking, based on analyses of persons with complete data on the predictor variables included in each model (Table 2). Maternal smoking and 7 of the 10 FAI variables were associated with increased odds of missing data on wheeze at age 81 months. Fewer variables were predictive of missing data

**Table 2.** Associations of Wheeze at Age 81 Months (Outcome Variable) and Prognostic and Family Adversity Index Variables With Missingness in the Outcome and Prognostic Variables, Avon Longitudinal Study of Parents and Children, Avon, United Kingdom, 1991–1999

| | Whole Data Set | | | Children With 4 Key Variables Measured[a] | | Odds of Missingness in Key Variables[b] | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | Wheeze at Age 81 Months (Outcome) | | Maternal Asthma | | Maternal Smoking | |
| | No. With Measurements | % With Measurements | % With Characteristic | No. With Measurements | % With Characteristic | OR | 95% CI | OR | 95% CI | OR | 95% CI |
| Wheeze at age 81 months | 8,402 | 60 | 13.4 | 8,037 | 13.3 | | | 1.34 | 0.98, 1.82 | 1.70 | 1.12, 2.59 |
| Male gender | 13,983 | 100 | 51.6 | 8,037 | 51.4 | 0.98 | 0.90, 1.07 | 0.68 | 0.46, 1.00 | 1.11 | 0.73, 1.70 |
| Maternal asthma | 12,303 | 88 | 11.6 | 8,037 | 11.3 | 1.05 | 0.91, 1.20 | | | 0.96 | 0.50, 1.87 |
| Maternal smoking | 13,163 | 94 | 19.9 | 8,037 | 15.4 | 1.44 | 1.29, 1.61 | 0.69 | 0.41, 1.17 | | |
| FAI variables measured during pregnancy | | | | | | | | | | | |
| Early parenthood | 13,983 | 100 | 8.1 | 8,037 | 4.7 | 1.87 | 1.57, 2.22 | 1.14 | 0.57, 2.27 | 1.13 | 0.51, 2.53 |
| Housing inadequacy | 13,399 | 96 | 7.4 | 7,992 | 4.7 | 1.54 | 1.28, 1.85 | 1.50 | 0.78, 2.90 | 0.42 | 0.13, 1.43 |
| Basic living conditions | 13,087 | 94 | 3.0 | 7,928 | 2.6 | 1.08 | 0.84, 1.40 | 0.81 | 0.25, 2.60 | 1.59 | 0.58, 4.39 |
| Low educational attainment | 12,503 | 89 | 14.2 | 7,927 | 11.4 | 1.30 | 1.15, 1.48 | 1.23 | 0.74, 2.04 | 1.93 | 1.15, 3.23 |
| Financial difficulties | 12,094 | 86 | 10.0 | 7,736 | 7.9 | 1.28 | 1.11, 1.49 | 1.24 | 0.71, 2.17 | 1.93 | 1.07, 3.50 |
| Large family size | 13,222 | 95 | 1.6 | 7,998 | 1.2 | 1.13 | 0.76, 1.68 | 0.36 | 0.05, 2.77 | 1.08 | 0.14, 8.44 |
| Affective psychopathology | 13,018 | 93 | 26.1 | 7,873 | 22.7 | 1.26 | 1.14, 1.39 | 1.84 | 1.22, 2.78 | 0.78 | 0.46, 1.30 |
| Crime trouble with police | 11,938 | 85 | 3.0 | 7,445 | 1.9 | 1.60 | 1.23, 2.07 | 0.60 | 0.18, 1.97 | —[c] | |
| No social support from partner | 12,086 | 86 | 13.0 | 7,884 | 10.7 | 1.09 | 0.95, 1.25 | 1.86 | 1.15, 2.99 | 1.18 | 0.64, 2.18 |
| Lack of social network | 12,414 | 89 | 7.6 | 8,021 | 5.6 | 1.51 | 1.28, 1.78 | 2.15 | 1.25, 3.68 | 1.17 | 0.55, 2.52 |

Abbreviations: CI, confidence interval; FAI, Family Adversity Index; OR, odds ratio.

[a] For FAI variables, number of children with 4 key variables and the FAI variable measured.

[b] Adjusted for other prognostic variables and, where applicable, the FAI variable.

[c] The association between "crime trouble with the police" and missingness of data on maternal smoking could not be estimated, because all women who smoked had missing data on that question.

**Table 3.** Associations of Prognostic Variables and Previous Measurements of Wheeze With Wheeze at Age 81 Months (Outcome Variable), Avon Longitudinal Study of Parents and Children, Avon, United Kingdom, 1991–1999

| | Whole Data Set | | | Children With 4 Key Variables Measured | | Odds Ratio for Wheeze at Age 81 Months[a] | 95% Confidence Interval |
|---|---|---|---|---|---|---|---|
| | No. With Measurements | % With Measurements | % With Characteristic | No. With Measurements | % With Characteristic | | |
| Male gender | 13,983 | 100 | 51.6 | 8,037 | 51.4 | 1.16 | 0.94, 1.44 |
| Maternal asthma | 12,303 | 88 | 11.6 | 8,037 | 11.3 | 1.63 | 1.22, 2.18 |
| Maternal smoking | 13,163 | 94 | 19.9 | 8,037 | 15.4 | 1.21 | 0.89, 1.65 |
| Previous measurement of wheeze at: | | | | | | | |
| 6 months | 11,409 | 82 | 26.4 | 7,738 | 24.3 | 1.12 | 0.88, 1.44 |
| 18 months | 10,976 | 78 | 27.5 | 7,706 | 26.2 | 1.20 | 0.93, 1.55 |
| 30 months | 9,990 | 71 | 22.6 | 7,391 | 21.4 | 1.37 | 1.05, 1.78 |
| 42 months | 10,004 | 72 | 17.6 | 7,579 | 16.7 | 2.12 | 1.62, 2.78 |
| 57 months | 9,411 | 67 | 18.9 | 7,532 | 18.0 | 3.14 | 2.43, 4.07 |
| 69 months | 8,600 | 62 | 15.5 | 7,339 | 15.0 | 14.10 | 11.2, 17.8 |

[a] Adjusted for other prognostic variables and, where applicable, the wheeze variable.

on maternal asthma or smoking during pregnancy. Wheeze at age 81 months was associated with missing data on both maternal asthma and smoking.

Table 3 shows the number of children with measurements and the percentage of those with the characteristic for the prognostic and wheeze variables. The number of children with measurements of wheeze declined from 11,409 at age 6 months to 8,402 at age 81 months. The prevalence of wheeze among children with the 4 key variables observed was lower than that among all those children who were observed at each time point. Wheeze at age 81 months was associated with maternal asthma, with weak evidence of associations with gender and maternal smoking. The strength of associations between earlier wheeze and wheeze at 81 months increased with age (for wheeze at age 69 months, adjusted odds ratio = 14.1, 95% confidence interval: 11.2, 17.8). This high odds ratio permits accurate prediction of wheeze at 81 months for those children ($n$ = 404) who are missing data on wheeze at 81 months and have observed data on wheeze at 69 months. There was little evidence of association between wheeze at 81 months and the FAI variables (odds ratios were 0.72–1.24; data not shown). There was also little evidence of association between earlier wheeze and missing data on the outcome or prognostic variables (odds ratios were 0.41–2.28; data not shown), although children with wheeze at 69 months were more likely to have missing data on wheeze at 81 months (odds ratio = 1.49, 95% confidence interval: 1.08, 2.06).

The associations examined in Table 2 (predictors of the probability of being missing) and Table 3 (predictors of values of variables with missing data) are summarized in Figure 1. Maternal smoking and wheeze at age 69 months were associated with both the outcome (wheeze at age 81 months) and the probability that it was missing. Therefore, it was not plausible to assume that the data were MCAR. If, however, given the (approximate) strata formed by these variables, we assume that the distributions of observed and missing 81-month wheeze data are similar, then 81-month wheeze is MAR. Under such an assumption, analyses will be unbiased, provided that both of these variables are included in complete-case analyses or that they are both included in the imputation models (24).

Table 4 shows estimates of the prevalence of wheeze at age 81 months and its association with the prognostic variables, based on both complete-case analyses and the 3 approaches to multiple imputation. All imputation models included gender, maternal asthma, maternal smoking, and 1) FAI variables only (model 1, predicting the probability of missingness); 2) wheeze variables only (model 2, predicting the values of variables with missing data); and 3) both FAI and wheeze variables (model 3, predicting both sets of variables). Assuming that the MAR assumption holds and that all models are correctly specified, the differences between imputed prevalence and association estimates and the corresponding complete-case estimates represent bias corrections. Under this assumption, the estimates from each

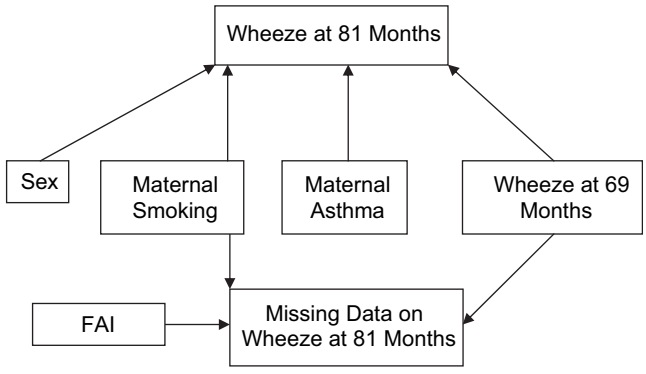


**Figure 1.** Associations between wheeze at age 81 months (outcome variable), covariates (gender, maternal smoking, and maternal asthma), and missingness of data, Avon Longitudinal Study of Parents and Children, Avon, United Kingdom, 1991–1999. FAI, Family Adversity Index.

**Table 4.** Prevalence and Associations of Wheeze at Age 81 Months With Prognostic Variables Based on Complete-Case Analyses and on 3 Approaches to Multiple Imputation, Avon Longitudinal Study of Parents and Children, Avon, United Kingdom, 1991–1999

| Analysis | Prevalence, % | 95% CI | SE of Prevalence | Estimated Fraction of Missing Information, % |
|---|---|---|---|---|
| Prevalence | | | | |
| Complete cases | 13.3 | 12.5, 14.0 | 0.38 | |
| Multiple imputation | | | | |
| FAI (predicts probability of missingness) | 13.7 | 12.9, 14.5 | 0.40 | 48.1 |
| Wheeze at ages 6–69 months (predicts values for missing data) | 14.1 | 13.4, 14.8 | 0.37 | 36.5 |
| Combined FAI and wheeze (predicts probability of missingness and values) | 14.1 | 13.3, 14.8 | 0.38 | 39.1 |
| | **OR** | **95% CI** | **SE of Log OR** | |
| Association with gender | | | | |
| Complete cases | 1.29 | 1.14, 1.47 | 0.066 | |
| Multiple imputation | | | | |
| FAI (predicts probability of missingness) | 1.32 | 1.16, 1.50 | 0.067 | 41.8 |
| Wheeze at ages 6–69 months (predicts values for missing data) | 1.39 | 1.23, 1.57 | 0.062 | 36.6 |
| Combined FAI and wheeze (predicts probability of missingness and values) | 1.38 | 1.22, 1.56 | 0.063 | 38.9 |
| Association with maternal asthma | | | | |
| Complete cases | 2.20 | 1.86, 2.61 | 0.087 | |
| Multiple imputation | | | | |
| FAI (predicts probability of missingness) | 2.23 | 1.87, 2.65 | 0.088 | 45.7 |
| Wheeze at ages 6–69 months (predicts values for missing data) | 2.24 | 1.91, 2.62 | 0.081 | 38.5 |
| Combined FAI and wheeze (predicts probability of missingness and values) | 2.26 | 1.92, 2.65 | 0.082 | 40.3 |
| Association with maternal smoking | | | | |
| Complete cases | 1.24 | 1.05, 1.47 | 0.087 | |
| Multiple imputation | | | | |
| FAI (predicts probability of missingness) | 1.25 | 1.06, 1.47 | 0.085 | 55.1 |
| Wheeze at ages 6–69 months (predicts values for missing data) | 1.32 | 1.14, 1.53 | 0.076 | 46.5 |
| Combined FAI and wheeze (predicts probability of missingness and values) | 1.32 | 1.14, 1.52 | 0.074 | 40.9 |

Abbreviations: CI, confidence interval; FAI, Family Adversity Index; OR, odds ratio; SE, standard error.

of the 3 imputation models show a bias correction in a consistent direction, with estimates from imputation models being higher than those from corresponding complete-case analyses. The largest change in estimates occurs when the wheeze variables (which predict the outcome) are included in the imputation model. Estimated prevalence increases

from 13.3% in the complete-case analysis to 14.1% using the full imputation model (including the prognostic, FAI, and wheeze variables, predicting both the probability of missingness and the values of variables with missing data). Using the full imputation model increases the odds ratios for the associations between wheeze at age 81 months and gender, maternal asthma, and maternal smoking by 7%, 3%, and 6%, respectively (Table 4) in comparison with complete-case analyses. There was little difference between results from the full imputation model (model 3) and the simpler model including only variables related to the value of the outcome (model 2).

The distributions of all variables were similar for observed and imputed data (Table 5), indicating no obvious problems with the imputation process.

Imputation including the wheeze variables (predicting values of variables with missing data) increased the efficiency of the analyses, assuming MAR holds (Table 4). The ratios of standard errors of log odds ratios comparing the imputation models including both wheeze and FAI variables with complete-case analyses were 0.96, 0.94, and 0.85 for the associations with gender, maternal asthma, and maternal smoking, respectively. These correspond to ratios of variances of 0.91, 0.89, and 0.78 and hence to hypothetical sample size increases of 10%, 12%, and 28%, respectively. The estimated fraction of missing information decreased from the complete-case analysis to the imputation model including the wheeze variables from age 6 months to age 69 months (predicting values of variables with missing data).

To quantify the variability due to the random sampling inherent in multiple imputation procedures, we calculated the standard deviations of 50 estimates, each from 1 multiple imputation analysis (model 3, predicting both the probability of missingness and the values of variables with missing data), varying the number of imputed data sets on which these estimates were based (50 estimates each from 5, 10, 20, 40, and 200 imputed data sets (Table 6)). When only 5 imputations were used, the standard deviations of the 50 imputation-based estimates of each association (Table 6) were approximately one-quarter of the magnitude of the

**Table 5.** Distributions of Outcome, Prognostic, Wheeze, and Family Adversity Index Variables for Observed and Imputed Data Sets, Avon Longitudinal Study of Parents and Children, Avon, United Kingdom, 1991–1999

| | % With Characteristic | |
|---|---|---|
| | Observed Data | Imputed Data |
| Wheeze at age 81 months | 13.4 | 15.0 |
| Maternal asthma | 11.6 | 12.8 |
| Maternal smoking | 19.9 | 22.8 |
| Wheeze at: | | |
| 6 months | 26.3 | 29.4 |
| 18 months | 27.5 | 30.6 |
| 30 months | 22.6 | 24.9 |
| 42 months | 17.6 | 20.0 |
| 54 months | 18.9 | 20.9 |
| 69 months | 15.5 | 17.2 |
| FAI variables measured during pregnancy | | |
| Housing inadequacy | 7.4 | 9.7 |
| Basic living conditions | 3.0 | 3.3 |
| Low educational attainment | 14.2 | 18.2 |
| Financial difficulties | 10.0 | 12.3 |
| Large family size | 1.6 | 1.7 |
| Affective psychopathology | 26.1 | 29.2 |
| Crime trouble with police | 3.0 | 4.1 |
| No social support from partner | 13.0 | 17.0 |
| Lack of social network | 7.6 | 10.2 |

Abbreviation: FAI, Family Adversity Index.

**Table 6.** Between-Imputation-Procedure Standard Deviations for 50 Sets of Imputations Under the Full Imputation Model, for Varying Numbers of Imputed Data Sets per Imputation Procedure, Avon Longitudinal Study of Parents and Children, Avon, United Kingdom, 1991–1999

| | No. of Imputations | | | | |
|---|---|---|---|---|---|
| | 5 | 10 | 20 | 40 | 200 |
| Prevalence (proportion) of wheeze at age 81 months (SE[a], 0.38) | 0.0884 | 0.0564 | 0.0458 | 0.0372 | 0.0170 |
| Log odds ratio for association between: | | | | | |
| Gender and wheeze at 81 months (SE, 0.063) | 0.0162 | 0.0099 | 0.0078 | 0.0050 | 0.0026 |
| Maternal asthma and wheeze at 81 months (SE, 0.082) | 0.0242 | 0.0162 | 0.0123 | 0.0093 | 0.0038 |
| Maternal smoking and wheeze at 81 months (SE, 0.074) | 0.0227 | 0.0186 | 0.0117 | 0.0097 | 0.0041 |

Abbreviation: SE, standard error.
[a] Standard error of estimate.

standard error from the set of 200 imputations. When 40 imputations were used, the ratio fell to approximately one-tenth.

## DISCUSSION

Multiple imputation has the potential to reduce bias and increase efficiency (reduce standard errors) in analyses of epidemiologic data, compared with complete-case analyses (3). However, its ability to reduce bias in a particular analysis depends on the existence of measured variables that are associated with both missingness and the outcome variable (7, 29–31). Preliminary analyses of such associations, and graphical displays summarizing them, can be used to investigate the plausibility of the assumptions underlying both complete-case and multiple-imputation-based analyses. Multiple-imputation-based analyses will not always reduce bias in comparison with complete-case analyses. Preliminary analyses thus inform the choice of analysis method and of variables to be included in any imputation model. Analyses using multiple imputation should often be based on 25 or more imputed data sets rather than the 3 or 5 that are often used, in order to reduce the impact of the random sampling inherent in multiple imputation procedures.

Changes in estimates were greater for prevalence than for association, with the complete-case analysis underestimating the prevalence of wheeze at age 81 months. Changes in odds ratios were not substantial but suggested that the complete-case analyses underestimated the associations between gender, maternal asthma, and maternal smoking and wheeze at 81 months. Marginal means (such as prevalence) are likely to show greater bias in complete-case analysis than estimates of association, because marginal means are always biased if data are MAR (rather than MCAR), whereas conditional associations may not be (if we condition on the MAR mechanism variables, as we did here). Bias corrections in estimates of association may be more substantial than those seen here when there are more measured variables that are strongly related to both missingness and the values of variables with missing data (30). Conversely, if most variables were weakly correlated with the variables with missing data, there would be little information to recover, and thus imputation would not substantially reduce bias. In our example, including the variables related to the values of the variable with the most missing data made the most difference in the estimates: Estimates from the full imputation model (additionally including variables related only to the probability of missingness) made little additional difference in the estimates.

Here, the outcome variable was the main variable with missing data. When only the outcome is missing, complete-case analysis is unbiased provided that missingness is unrelated to the outcome variable, given the covariates (6, 7). Our preliminary analyses (Figure 1) showed that wheeze at age 69 months and maternal smoking were related to both the outcome (wheeze at age 81 months) and the probability of the outcome being missing. If the analysis model included both of these variables and the outcome was MAR given these variables, we would expect complete-case analysis to be unbiased. Because wheeze at age 69 months was not included as a covariate in the complete-case analyses, the complete-case analyses will be biased. However, if in Figure 1 wheeze at age 69 months was unrelated to the probability of the outcome being missing, given maternal smoking status, then a complete-case analysis with wheeze at age 81 months as the outcome and maternal smoking as a covariate would be unbiased (7).

Where there are nontrivial amounts of missing data in covariates, both preliminary analyses and imputation models will become more complex. An MAR assumption may often become more plausible after the inclusion in the imputation model of additional variables that are not in our analysis model (because they are on the causal pathway, for example). Thus, multiple imputation models should typically be more complex than the analysis model. Including variables that are not related to the variable being imputed in the imputation models may slightly decrease efficiency but should not cause bias (29, 31). Model diagnostics should be used to highlight any implausibility in the imputed values. For example, the distributions of observed and imputed data should be compared and the plausibility of any differences examined. Imputation models should also preserve the structure of the analysis model (32). For example, where the substantive analysis exploits the hierarchical nature of longitudinal data (e.g., using a multilevel model), the imputation model should be similarly structured. Here, the longitudinal nature of the data allowed us to include variables (previous wheezing) that predicted the values of the variable with the most missing data (wheeze at 81 months) in imputation models.

As well as correcting bias, multiple imputation will often improve efficiency compared with complete-case analyses. Here, in estimating both prevalence and associations, the standard errors from complete-case analyses were larger than those from the full imputation model, with improvements in efficiency corresponding to hypothetical sample size increases of up to 28% relative to complete-case analyses, for the association between maternal asthma and wheeze at age 81 months. Inclusion in the imputation model of variables that are strongly related to the variable with missing data (even if unrelated to missingness) will usually improve efficiency (31). Here, the variable most related to wheeze at age 81 months was wheeze at age 69 months. The standard errors for estimates where the imputation model included wheeze at 69 months were lower in all cases than those from the complete-case analyses and those where the imputation model did not include wheeze at 69 months.

When there are factors related to both the outcome and missingness that are not included in imputation models, the data are MNAR and multiple imputation using standard procedures cannot (fully) remove bias. For example, the data analyzed here would be MNAR if mothers of children with wheeze at 81 months were less likely to return the questionnaire, even after allowing for all measured variables. Our preliminary analyses allow some insight into the direction of the association between missingness and wheeze—for example, children who had wheeze at 69 months were more likely to have missing data on wheeze

at 81 months, after allowing for other variables such as maternal smoking status. Given the strong positive association also shown here between wheeze at 69 months and wheeze at 81 months, this suggests that mothers of children with wheeze at 81 months may be less likely to complete questionnaires about wheeze and thus that the true prevalence of wheeze is higher than estimated here.

Once the multiple imputation model is chosen, the number of imputations must be decided. The variability between sets of imputations depends on both the number of imputations used and the fraction of missing information (27). However, the fraction of missing information is itself estimated using the between- and within-imputation variances, and thus may have substantial variability when estimated from small numbers of imputations. Monte Carlo variation among sets of small numbers of imputations can be substantial enough to materially affect conclusions, particularly where the original data set is small (14, 27). One approach might be to estimate the Monte Carlo variation (33) and use that to decide the appropriate number of imputations. For example, the desired precision of the estimate could be decided, a small number of imputations (20, for example) could be carried out to obtain a jackknife estimate of the Monte Carlo variance, and then the number of imputations required to achieve the desired precision could be calculated (33).

Multiple imputation is increasingly used, and it has been suggested that ''if correctly and thoughtfully applied, imputation methods should reduce bias and increase precision in everyday use'' (4, p. 356). However, correct application is not simple, and multiple imputation is not less biased than complete-case analysis in all circumstances (7). In this paper, we have shown that useful information for constructing the imputation model, on the likely extent of the bias correction, and on the potential efficiency gains from multiple imputation can be obtained from careful preliminary analyses. These should include exploration of the factors related to missingness and their association with variables in the analysis model, as well as exploration of variables predictive of variables in the analysis model (even if not predictive of missingness). We found that imputation including variables related to the values of the variable with the most missing data had the greatest impact on the estimates and their standard errors; additionally including variables related only to the probability of missingness had little additional impact. In our example, preliminary analyses would have highlighted the importance of including earlier wheeze variables, while indicating that the FAI variables (being related only to the probability of missingness) need not be included. Thus, such analyses should be reported in papers using or considering multiple imputation, either as justification for the variables used in the multiple imputation (4) or as reasons for preferring complete-case analysis over multiple imputation.

## ACKNOWLEDGMENTS

Author affiliations: Department of Social Medicine, Faculty of Medicine and Dentistry, University of Bristol, Bristol, United Kingdom (Michael Spratt, Jonathan A. C. Sterne, Jon Heron, Kate Tilling); Medical Statistics Unit, London School of Hygiene and Tropical Medicine, University of London, London, United Kingdom (James Carpenter); Clinical Epidemiology and Biostatistics Unit, Murdoch Children's Research Institute, Royal Children's Hospital, Parkville, Victoria, Australia (John B. Carlin); Department of Paediatrics, Faculty of Medicine, Dentistry and Health Sciences, University of Melbourne, Royal Children's Hospital, Parkville, Victoria, Australia (John B. Carlin); and Department of Community Based Medicine, Faculty of Medicine and Dentistry, University of Bristol, Bristol, United Kingdom (John Henderson).

This work was supported by the United Kingdom Medical Research Council (grant G0600599) and the Economic and Social Research Council (grant RES-063-27-0257). The Medical Research Council, the Wellcome Trust, and the University of Bristol provide core support for the Avon Longitudinal Study of Parents and Children.

The authors are grateful to the midwives for their help in recruiting study participants and to the entire study team, which includes interviewers, computer and laboratory technicians, clerical workers, research scientists, volunteers, managers, receptionists, and nurses.

This publication is the work of the authors, and Dr. Kate Tilling will serve as guarantor for its contents.

Conflict of interest: none declared.

## REFERENCES

1. Diggle PJ, Heagerty P, Liang K-Y, et al. *Analysis of Longitudinal Data*. Oxford, United Kingdom: Oxford University Press; 2002.
2. Verbeke G, Molenberghs G. *Linear Mixed Models for Longitudinal Data*. New York, NY: Springer-Verlag New York, Inc; 2000.
3. Sterne JA, White IR, Carlin JB, et al. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. *BMJ*. 2009;338:b2393. (doi: 10.1136/bmj.b2393).
4. Klebanoff MA, Cole SR. Use of multiple imputation in the epidemiologic literature. *Am J Epidemiol*. 2008;168(4):355–357.
5. Rubin DB. Inference and missing data. *Biometrika*. 1976;63(3):581–590.
6. Little RJA, Rubin DB. *Statistical Analysis With Missing Data*. 2nd ed. Hoboken, NJ: John Wiley & Sons, Inc; 2002.
7. White IR, Carlin JB. Bias and efficiency of multiple imputation compared with complete-case analysis for missing covariate values. *Stat Med*. In press.
8. Steyerberg EW, van Veen M. Imputation is beneficial for handling missing data in predictive models [letter]. *J Clin Epidemiol*. 2007;60(9):979.
9. Allison PD. Multiple imputation for missing data—a cautionary tale. *Sociol Methods Res.* 2000;28(3):301–309.
10. Schafer JL. *Analysis of Incomplete Multivariate Data*. 1st ed. London, United Kingdom: Chapman & Hall Ltd; 1997.
11. Allison PD. *Missing Data*. Thousand Oaks, CA: Sage Publications, Inc; 2001.
12. van Buuren S, Boshuizen HC, Knook DL. Multiple imputation of missing blood pressure covariates in survival analysis. *Stat Med*. 1999;18(6):681–694.

13. Rubin DB. Multiple imputation after 18+ years. *J Am Stat Assoc.* 1996;91(434):473–489.
14. Royston P. Multiple imputation of missing values: update. *Stata J*. 2005;5(2):188–201.
15. Carlin JB, Li N, Greenwood P, et al. Tools for analysing multiple imputed datasets. *Stata J*. 2003;3(3):1–20.
16. Schafer J. *NORM: Multiple Imputation of Incomplete Multivariate Data Under a Normal Model*. University Park, PA: Department of Statistics, Pennsylvania State University; 1997.
17. Royston P. Multiple imputation of missing values. *Stata J*. 2004;4(3):227–241.
18. Raghunathan TE, Solenberger PW, Van Hoewyk J. *IVEware: Imputation and Variance Estimation Software. User Guide*. Ann Arbor, MI: Institute for Social Research, University of Michigan; 2002.
19. Horton NJ, Lipsitz SR. Multiple imputation in practice: comparison of software packages for regression models with missing variables. *Am Stat*. 2001;55(3):244–254.
20. Statistical Solutions Ltd. *SOLAS 3.0*. Cork, Ireland: Statistical Solutions Ltd: 2004.
21. Schafer JL. Multiple imputation: a primer. *Stat Methods Med Res.* 1999;8(1):3–15.
22. Graham JW, Olchowski AE, Gilreath TD. How many imputations are really needed? Some practical clarifications of multiple imputation theory. *Prev Sci.* 2007;8(3):206–213.
23. Golding J, Pembrey M, Jones R, et al. ALSPAC—the Avon Longitudinal Study of Parents and Children. I. Study methodology. *Paediatr Perinat Epidemiol*. 2001;15(1):74–87.
24. Nemes S, Jonasson JM, Genell A, et al. Bias in odds ratios by logistic regression modelling and sample size. *BMC Med Res Methodol*. 2009;9:56. (doi: 10.1186/1471-2288-9-56).
25. Butler RJ, Heron J. The prevalence of infrequent bedwetting and nocturnal enuresis in childhood. A large British cohort. *Scand J Urol Nephrol*. 2008;42(3):257–264.
26. Steer C, Wolke D, ALSPAC Study Team. An index of family adversity [poster]. Presented at the 3rd Conference on Epidemiological Longitudinal Studies in Europe, Bristol, United Kingdom, September 22–24, 2004.
27. Bodner TE. What improves with increased missing data imputations? *Struct Equ Modeling*. 2008;15(4):651–75.
28. Stata Corporation. *Stata Statistical Software: Release 10*. College Station, TX: Stata Corporation; 2007.
29. Kenward MG, Carpenter J. Multiple imputation: current perspectives. *Stat Methods Med Res.* 2007;16(3):199–218.
30. Moons KG, Donders RA, Stijnen T, et al. Using the outcome for imputation of missing predictor values was preferred. *J Clin Epidemiol*. 2006;59(10):1092–1101.
31. Collins LM, Schafer JL, Kam CM. A comparison of inclusive and restrictive strategies in modern missing data procedures. *Psychol Methods*. 2001;6(4):330–351.
32. Harel O, Zhou XH. Multiple imputation: review of theory, implementation and software. *Stat Med*. 2007;26(16):3057–3077.
33. Royston P, Carlin JB, White IR. Multiple imputation of missing values: new features for mim. *Stata J*. 2009;9(2):252–264.