Biometrics 68, 129–137
March 2012
DOI: 10.1111/j.1541-0420.2011.01666.x

Combining Multiple Imputation and Inverse-Probability Weighting

Shaun R. Seaman,[1,*] Ian R. White,[1] Andrew J. Copas,[2,3] and Leah Li[4]

[1]MRC Biostatistics Unit, Cambridge, CB2 0SR, U.K.
[2]MRC Clinical Trials Unit, London, NW1 2DA, U.K.
[3]Centre for Sexual Health and HIV Research, University College London, London, WC1E 6JB, U.K.
[4]MRC Centre of Epidemiology for Child Health, UCL Institute of Child Health, London, WC1N 1EH, U.K.
*email: shaun.seaman@mrc-bsu.cam.ac.uk

Summary. Two approaches commonly used to deal with missing data are multiple imputation (MI) and inverse-probability weighting (IPW). IPW is also used to adjust for unequal sampling fractions. MI is generally more efficient than IPW but more complex. Whereas IPW requires only a model for the probability that an individual has complete data (a univariate outcome), MI needs a model for the joint distribution of the missing data (a multivariate outcome) given the observed data. Inadequacies in either model may lead to important bias if large amounts of data are missing. A third approach combines MI and IPW to give a doubly robust estimator. A fourth approach (IPW/MI) combines MI and IPW but, unlike doubly robust methods, imputes only isolated missing values and uses weights to account for remaining larger blocks of unimputed missing data, such as would arise, e.g., in a cohort study subject to sample attrition, and/or unequal sampling fractions. In this article, we examine the performance, in terms of bias and efficiency, of IPW/MI relative to MI and IPW alone and investigate whether the Rubin's rules variance estimator is valid for IPW/MI. We prove that the Rubin's rules variance estimator is valid for IPW/MI for linear regression with an imputed outcome, we present simulations supporting the use of this variance estimator in more general settings, and we demonstrate that IPW/MI can have advantages over alternatives. IPW/MI is applied to data from the National Child Development Study.

Key words: Marginal model; Missing at random; Survey weighting; 1958 British Birth Cohort.

1. Introduction

Datasets collected for medical or social research contain missing values. One approach for dealing with this problem is simply to exclude individuals with missing data. This "complete-case" analysis is valid when data are missing completely at random but not necessarily when missing at random (MAR) (Little and Rubin, 2002). It can also be inefficient. Two alternatives are inverse-probability weighting (IPW) (Höfler et al., 2005) and multiple imputation (MI) (Little and Rubin, 2002). In IPW, again only complete cases are included in the analysis (excepting analysis of repeated measures, which we do not treat here), but weights are used to rebalance the set of complete cases so that it is representative of the whole sample. Inverse-probability weights can also be used to adjust for different sampling fractions in a survey. They are then known as sampling weights and rebalance the sample to make it representative of the population.

In MI, missing data are replaced by data drawn from an imputation model. This is done M times, generating M complete datasets. Each is analyzed and an estimate of the model parameters, $\boldsymbol{\theta}$, calculated. Let $\widehat{\boldsymbol{\theta}}$ denote the complete-data estimator of $\boldsymbol{\theta}$, and $\widehat{\boldsymbol{V}}$ its estimated variance. Let $\widehat{\boldsymbol{\theta}}_{(m)}$ and $\widehat{\boldsymbol{V}}_{(m)}$ be their values for the mth imputed dataset ($m = 1, \ldots, M$).

Rubin (1987) proposed $\boldsymbol{\theta}$ be estimated by $\widehat{\boldsymbol{\theta}}_M$ and $\mathrm{Var}(\widehat{\boldsymbol{\theta}}_M)$ by $\widehat{\boldsymbol{V}}_M$, where

$$\widehat{\boldsymbol{\theta}}_M = \frac{1}{M}\sum_{m=1}^{M}\widehat{\boldsymbol{\theta}}_{(m)}, \quad (1)$$

$$\widehat{\boldsymbol{V}}_M = \frac{1}{M}\sum_{m=1}^{M}\widehat{\boldsymbol{V}}_{(m)} + (1+M^{-1})(M-1)^{-1} \times \sum_{m=1}^{M}(\widehat{\boldsymbol{\theta}}_{(m)} - \widehat{\boldsymbol{\theta}}_M)(\widehat{\boldsymbol{\theta}}_{(m)} - \widehat{\boldsymbol{\theta}}_M)^T. \quad (2)$$

IPW and MI yield consistent estimators of $\boldsymbol{\theta}$ when the data are MAR and the imputation and weighting models, respectively, are correctly specified. The variance of the IPW estimator is consistently estimated provided the weighting is taken into account, e.g., using a sandwich estimator (Robins, Rotnitzky, and Zhao, 1994). For MI, when $\widehat{\boldsymbol{\theta}}$ is the maximum likelihood estimator (MLE), $\widehat{\boldsymbol{V}}$ is the inverse Fisher information, and missing data are sampled from their Bayesian posterior predictive distribution, $\widehat{\boldsymbol{\theta}}_M$ is asymptotically normally distributed with variance $\boldsymbol{V}_M$, and $\widehat{\boldsymbol{V}}_M$ is an asymptotically unbiased estimator of $\boldsymbol{V}_M$ and is consistent when $M = \infty$ (Rubin, 1987; Wang and Robins, 1998; Nielsen, 2003).

MI is often preferred to IPW, as it is usually more efficient. If the imputation model is correctly specified, MI should work well. However, if many data are being imputed, any inadequacies in the imputation model may lead to considerable bias. If few variables are missing on an individual, it may be considered desirable to impute them, rather than exclude the individual. On the other hand, if many variables are missing on the same individual, the imputation model must describe the joint distribution of all these variables, and if many individuals have many missing variables, the analyst may be nervous about relying on this complex and possibly misspecified imputation model. This situation could arise, for example, in a longitudinal study when whole blocks of data are missing on some of the individuals due to missed visits, or in a survey when some individuals have declined to answer whole sets of related questions. In such situations, the analyst may feel more confident using IPW.

Another possibility is to combine MI and IPW. A rule is specified for when to include an individual in the analysis: e.g., if they attended a follow-up visit, or if more than a certain percentage of their data is observed. Missing values in included individuals are multiply imputed and each resulting dataset (which we call a "quasi-complete dataset" because the data are complete for the included, but not excluded, individuals) is analyzed using IPW to account for the exclusion of individuals not satisfying the inclusion rule and for different sampling fractions (if any). The "quasi-complete-data" estimator $\widehat{\boldsymbol{\theta}}$ is then the IPW estimator using the data on included individuals in a single quasi-complete dataset and $\widehat{\boldsymbol{V}}$ is the corresponding sandwich variance estimator. We call this method "IPW/MI." By imputing in individuals with few missing values but excluding individuals with more missing data, IPW/MI could inherit some of the efficiency advantage of MI while avoiding bias resulting from incorrectly imputing larger blocks of data. IPW/MI is also needed when sampling weights are used together with MI, even if all individuals are included in the analysis.

Several authors have used IPW/MI. Caldwell et al. (2008) and Stansfeld et al. (2008a,2008b) analyzed data from the National Childhood Development Study (NCDS). They regressed outcomes measured at age 45 on predictors measured at the same or earlier visits. Attrition of the cohort over time meant that 41% missed the age 45 visit. Weights were used to adjust for attrition, while missing values in those who attended the visit were multiply imputed. Priebe et al. (2004) multiply imputed missing data in a logistic regression with sampling weights.

It is not obvious that Rubin's rules will give valid variance estimators for IPW/MI. IPW estimators are inefficient. Robins and Wang (2000) and Nielsen (2003) show for MI that when $\widehat{\boldsymbol{\theta}}$ is inefficient, $\widehat{\boldsymbol{V}}_M$ can be asymptotically biased, even if $\widehat{\boldsymbol{V}}$ is a consistent estimator of the complete-data variance and imputation is from the correct posterior predictive distribution. The purpose of the present article is twofold: to examine asymptotic bias in $\widehat{\boldsymbol{V}}_M$ when $\widehat{\boldsymbol{\theta}}$ is an IPW estimator and to show when IPW/MI is useful.

In Section 2, we define IPW/MI and show it gives consistent estimation of $\boldsymbol{\theta}$. In Section 3, we show $\widehat{\boldsymbol{V}}_M$ is asymptotically unbiased for IPW/MI with linear regression and imputed outcomes. Section 4 describes a simulation study verifying this and demonstrating IPW/MI can have advantages over MI or IPW alone. Section 5 is a simulation with imputed covariate, suggesting $\widehat{\boldsymbol{V}}_M$ is approximately unbiased in this case. Section 6 is an application to NCDS.

2. IPW/MI and Consistency of $\widehat{\boldsymbol{\theta}}$

In this section, we describe IPW/MI for the situation where there are no sampling weights. The inclusion of sampling weights is covered in the Web Appendix available online.

An independent random sample of size N is drawn from the population. Let $\boldsymbol{D}$ denote, for an individual, the vector of the set of variables included in the analysis model as well as possibly other variables that will be used to impute missing values in that set of variables. Let R denote the missingness pattern in $\boldsymbol{D}$ (i.e., which elements of $\boldsymbol{D}$ are missing), and write $\boldsymbol{D} = (\boldsymbol{D}^o, \boldsymbol{D}^m)$, where $\boldsymbol{D}^o$ and $\boldsymbol{D}^m$ denote the observed and missing parts of $\boldsymbol{D}$, respectively. Subscript i denotes individual i in the sample; e.g., $\boldsymbol{D}_i$ denotes $\boldsymbol{D}$ for individual i.

The IPW/MI method is as follows. Let $\mathcal{R}(R)$ be a binary function of R chosen by the analyst. $\mathcal{R}(R)$ is the rule determining whether an individual is included in the analysis. An example of $\mathcal{R}(R)$ is $\mathcal{R}(R) = 1$ if fewer than a certain percentage of variables in the analysis model are missing and $\mathcal{R}(R) = 0$ otherwise. Let $\mathcal{A}$ denote the set of indices of individuals with $\mathcal{R}(R) = 1$. As formalized below, we estimate $\boldsymbol{\theta}$ by fitting the analysis model only to individuals $i \in \mathcal{A}$, using inverse-probability weights to account for the selection by $\mathcal{R}(R)$. Missing values in individuals $i \in \mathcal{A}$ are multiply imputed.

To impute $\boldsymbol{D}_i^m$ in individuals $i \in \mathcal{A}$, we assume a model $g(\boldsymbol{D}; \boldsymbol{\psi})$ for the conditional distribution of $\boldsymbol{D}$ given $\mathcal{R}(R) = 1$ with parameters $\boldsymbol{\psi}$. We say this model is correctly specified if $\exists\, \boldsymbol{\psi}_0$ such that $g(\boldsymbol{D}; \boldsymbol{\psi}_0)$ is the true distribution of $\boldsymbol{D}$ given $\mathcal{R}(R) = 1$. $\boldsymbol{\psi}$ is estimated by $\widehat{\boldsymbol{\psi}}$, its MLE using only the data on individuals $i \in \mathcal{A}$. Imputation may be proper or improper. Let $\boldsymbol{D}^{(m)}$ denote the mth imputed value of $\boldsymbol{D}$ ($m = 1, \ldots, M$). Note that if some elements of $\boldsymbol{D}$ are observed in all individuals with $\mathcal{R}(R) = 1$, the imputation model can be a model for the distribution of the remaining elements of $\boldsymbol{D}$ given these elements and $\mathcal{R}(R) = 1$.

Let $\boldsymbol{H}$ be a vector of fully observed variables that predict whether $\mathcal{R}(R) = 1$. Assume a model $w(\boldsymbol{H}; \boldsymbol{\alpha})$ for $P\{\mathcal{R}(R) = 1 \mid \boldsymbol{H}\}^{-1}$, where $\boldsymbol{\alpha}$ are parameters. We say this model is correctly specified if $\exists \boldsymbol{\alpha}_0$ such that $P\{\mathcal{R}(R) = 1 \mid \boldsymbol{H} = \boldsymbol{h}\} = w(\boldsymbol{h}; \boldsymbol{\alpha}_0)^{-1}\ \forall\, \boldsymbol{h}$. Let $W = w(\boldsymbol{H}; \boldsymbol{\alpha}_0)$. Assume $\exists\, \delta > 0$ such that $P(W^{-1} > \delta) = 1$. Typically, $\boldsymbol{\alpha}_0$, the true value of $\boldsymbol{\alpha}$, will be unknown. Let $\widehat{\boldsymbol{\alpha}}$ equal $\boldsymbol{\alpha}_0$ if $\boldsymbol{\alpha}_0$ is known and denote the MLE of $\boldsymbol{\alpha}$ otherwise.

Let $S_{\boldsymbol{\theta}}(\boldsymbol{\theta}; \boldsymbol{D})$ denote an individual's contribution to the (unweighted) complete-data estimating equations of the analysis model. Let $\boldsymbol{\theta}_0$ denote the solution of $E_{\boldsymbol{D}}\{S_{\boldsymbol{\theta}}(\boldsymbol{\theta}; \boldsymbol{D})\} = \boldsymbol{0}$. Therefore, $\boldsymbol{\theta}_0$ is the "true" value of $\boldsymbol{\theta}$: it is the value to which the solution to estimating equations $\sum_{i=1}^{N} S_{\boldsymbol{\theta}}(\boldsymbol{\theta}; \boldsymbol{D}_i) = \boldsymbol{0}$ would converge as $N \to \infty$. Based just on data from individuals $i \in \mathcal{A}$, let $\widehat{\boldsymbol{\theta}}_{(m)}$ be the solution to (weighted) estimating equations $\sum_{i \in \mathcal{A}} w(\boldsymbol{H}_i; \widehat{\boldsymbol{\alpha}}) \boldsymbol{S}_{\boldsymbol{\theta}}(\widehat{\boldsymbol{\theta}}_{(m)}; \boldsymbol{D}_i^{(m)}) = \boldsymbol{0}$ and let $\widehat{\boldsymbol{\theta}}_M$ be given by equation (1). Theorem 1 and its corollary state that under specified conditions $\widehat{\boldsymbol{\theta}}_M$ is a consistent estimator of $\boldsymbol{\theta}$. Proofs are given in the Web Appendix.

THEOREM 1. *Assume (i) model $w(\boldsymbol{H};\boldsymbol{\alpha})$ is correctly specified, (ii) $g(\boldsymbol{D};\boldsymbol{\psi})$ is correctly specified, (iii) $P\{\mathcal{R}(R)=1 \mid \boldsymbol{D},\boldsymbol{H}\} = P\{\mathcal{R}(R)=1 \mid \boldsymbol{H}\}$, (iv) $p\{R \mid \boldsymbol{D}, W, \mathcal{R}(R)=1\} = p\{R \mid \boldsymbol{D}^o, W, \mathcal{R}(R)=1\}$, and (v) $\boldsymbol{D}^m \perp\!\!\!\perp W \mid \boldsymbol{D}^o, \mathcal{R}(R)=1$. Then, when $M=\infty, \widehat{\boldsymbol{\theta}}_M \to \boldsymbol{\theta}_0$ as $N \to \infty$.*

Condition (iii) states that the probability an individual is used in the fitting of the imputation and analysis models does not depend on his values of the variables ($\boldsymbol{D}$) used in those models given the covariates ($\boldsymbol{H}$) in the weighting model. Condition (iv) states that among individuals to whom the imputation model is fitted, $\boldsymbol{D}$ is MAR given the true weight W. Condition (v) adds to this that among these individuals the missing variables in the imputation model must be conditionally independent of W given the observed variables. Note condition (v) can be satisfied by including W or $\boldsymbol{H}$ in $\boldsymbol{D}$. The necessity for condition (v) can be understood by considering how imputation will work if it is not satisfied. Set $\mathcal{A}$ is enriched for individuals with small values of W (and contains fewer with large values) compared to the entire sample. If (v) is false, the distribution of $\boldsymbol{D}$ given W depends on W, and when the imputation model is fitted to set $\mathcal{A}$, the resulting estimate of the marginal distribution of $\boldsymbol{D}$ will be biased toward the conditional distribution of $\boldsymbol{D}$ given small values of W. Missing data in all individuals in $\mathcal{A}$ are then imputed using the same model, a model that has been estimated giving too much weight to individuals with small W. Including W (or $\boldsymbol{H}$) in the imputation model avoids this problem: individuals with different W are imputed differently.

The following corollary shows that an alternative to including the true weights ($W = w(\boldsymbol{H};\boldsymbol{\alpha}_0)$) or the covariates that predict the weights ($\boldsymbol{H}$) in the imputation model is to include the estimated weights ($w(\boldsymbol{H};\widehat{\boldsymbol{\alpha}})$). The latter may be appealing because true weights are typically unknown and the dimension of $\boldsymbol{H}$ may be large.

COROLLARY 1. *Suppose the imputation model includes, in addition to $\boldsymbol{D}, w(\boldsymbol{H};\boldsymbol{\alpha})$. Assume conditions (i), (iii), and (iv) of Theorem 1 are satisfied, the imputation model $g\{\boldsymbol{D}, w(\boldsymbol{H};\boldsymbol{\alpha}_0);\boldsymbol{\psi}\}$ is correctly specified, $\boldsymbol{\psi}$ is estimated by its MLE $\widehat{\boldsymbol{\psi}}$ at $\boldsymbol{\alpha}=\widehat{\boldsymbol{\alpha}}$ using only individuals $i \in \mathcal{A}$, and $\boldsymbol{D}^m$ is imputed using $g\{\boldsymbol{D}, w(\boldsymbol{H};\widehat{\boldsymbol{\alpha}});\widehat{\boldsymbol{\psi}}\}$. Then, when $M=\infty, \widehat{\boldsymbol{\theta}}_M \to \boldsymbol{\theta}_0$ as $N \to \infty$.*

3. Linear Regression with Imputed Outcome

Consider the special case of linear regression with an imputed outcome. As in Section 2, we assume that there are no sampling weights; the generalization to sampling weights is given in the Web Appendix. Write $\boldsymbol{D} = (\boldsymbol{X}, Y)$ and let $\boldsymbol{Z}$ be $\boldsymbol{X}$ or a subvector of $\boldsymbol{X}$. Below, Y and $\boldsymbol{Z}$ will be the response and covariates, respectively, in the analysis model. Let $\mathcal{R}(R)=1$ if $\boldsymbol{X}$ is complete; $\mathcal{R}(R)=0$ otherwise. Let $R_Y=1$ if $\mathcal{R}(R)=1$ and Y is observed; $R_Y=0$ otherwise. We assume weights W are known and $\exists\, \delta > 0$ such that $P(W^{-1} > \delta) = 1$.

We estimate $\boldsymbol{\theta}$ in the analysis model

$$Y = \boldsymbol{\theta}^T \boldsymbol{Z} + e, \qquad \text{where } E(e \mid \boldsymbol{Z}) = 0. \tag{3}$$

by linear regression of Y on $\boldsymbol{Z}$. Therefore, $\boldsymbol{S}_\theta(\boldsymbol{\theta};\boldsymbol{D}) = \boldsymbol{Z}(Y - \boldsymbol{Z}^T\boldsymbol{\theta})$. The true value of $\boldsymbol{\theta}$ is the solution of $E_{\boldsymbol{D}}\{S_{\boldsymbol{\theta}}(\boldsymbol{\theta};\boldsymbol{D})\} = \boldsymbol{0}$, which is $\boldsymbol{\theta}_0 = \{E(\boldsymbol{Z}\boldsymbol{Z}^T)\}^{-1}E(\boldsymbol{Z}Y)$. We say the analysis model is correctly specified if equation (3) holds $\forall \boldsymbol{Z}$ when $\boldsymbol{\theta}=\boldsymbol{\theta}_0$; otherwise it is misspecified.

The quasi-complete-data estimator, $\widehat{\boldsymbol{\theta}}$, is the solution to weighted estimating equations $\sum_{i\in\mathcal{A}} W_i \boldsymbol{S}_{\boldsymbol{\theta}}(\widehat{\boldsymbol{\theta}};\boldsymbol{D}_i) = \boldsymbol{0}$, which is the weighted least squares estimator $\widehat{\boldsymbol{\theta}} = (\sum_{i\in\mathcal{A}} W_i \boldsymbol{Z}_i \boldsymbol{Z}_i^T)^{-1} \sum_{i\in\mathcal{A}} W_i \boldsymbol{Z}_i Y_i$. The quasi-complete-data variance estimator $\widehat{\boldsymbol{V}}$ is the sandwich estimator $(\sum_{i\in\mathcal{A}} W_i \boldsymbol{Z}_i \boldsymbol{Z}_i^T)^{-1} \{\sum_{i\in\mathcal{A}} W_i^2 \boldsymbol{Z}_i \boldsymbol{Z}_i^T (Y_i - \hat{\boldsymbol{\theta}}^T \boldsymbol{Z}_i)^2\}(\sum_{i\in\mathcal{A}} W_i \boldsymbol{Z}_i \boldsymbol{Z}_i^T)^{-1}$. Missing Y values in individuals $i \in \mathcal{A}$ are multiply imputed using $\boldsymbol{X}$ as predictors, $\widehat{\boldsymbol{\theta}}$ and $\widehat{\boldsymbol{V}}$ are calculated for each imputed dataset, and $\widehat{\boldsymbol{\theta}}_M$ and $\widehat{\boldsymbol{V}}_M$ are calculated from equations (1) and (2).

THEOREM 2. *Let missing Y be imputed from their posterior predictive distributions using the regression imputation procedure of Schenker and Welsh (1988) (p. 1560) with imputation model*

$$\begin{aligned} &Y_i = \boldsymbol{\beta}^T \boldsymbol{X}_i + \epsilon_i \quad (i \in \mathcal{A}), \\ &\textit{with } \{\epsilon_i : i \in \mathcal{A}\} \mid \{\boldsymbol{X}_i, W_i, \mathcal{R}(R_i)=1 : i \in \mathcal{A}\} \text{ i.i.d. } N\big(0, \sigma_\epsilon^2\big), \end{aligned} \tag{4}$$

and improper prior density for $(\boldsymbol{\beta}, \sigma_\epsilon^2)$ proportional to σ_ϵ^{-2}. Assume this model is correctly specified, i.e., there exists a $\boldsymbol{\beta}$ for which equation (4) holds, and that

$$P\{R_Y = 1 \mid Y, W, \boldsymbol{X}, \mathcal{R}(R)=1\} = P\{R_Y = 1 \mid W, \boldsymbol{X}, \mathcal{R}(R)=1\}. \tag{5}$$

Then (i) $\widehat{\boldsymbol{\theta}}_M$ is a consistent estimator of $\boldsymbol{\theta}$; (ii) if $\boldsymbol{X}$ includes $W\boldsymbol{Z}$ (i.e., $W_i\boldsymbol{Z}_i = \boldsymbol{C}\boldsymbol{X}_i\ \forall i$ for some matrix of constants $\boldsymbol{C}$), $\widehat{\boldsymbol{V}}_M$ is an asymptotically ($N \to \infty$) unbiased estimator of $\mathrm{Var}(\widehat{\boldsymbol{\theta}}_M)$; and iii) if $\boldsymbol{X}$ includes $W\boldsymbol{Z}$ and $M=\infty, \widehat{\boldsymbol{V}}_M$ is a consistent estimator of $\mathrm{Var}(\widehat{\boldsymbol{\theta}}_M)$.

Including $W\boldsymbol{Z}$ in $\boldsymbol{X}$ means including the pairwise interactions between the weight and all the variables in $\boldsymbol{Z}$, as well as (if the analysis model includes an intercept term) the weights themselves. Proofs of parts (i) and (ii) come from extending the proof of Kim et al. (2006), which shows (ii) is true in the special case where $\boldsymbol{Z}=1$; that of part (iii) comes from applying Theorem 2 of Robins and Wang (2000). Details are in the Web Appendix.

The reason $W\boldsymbol{Z}$ needs to be in $\boldsymbol{X}$ is to avoid the imputer assuming more than the analyst. Consider the simple case where $\boldsymbol{Z}=1$ (so θ is the population mean) and there are two values of W: a and b. The complete-data estimator of θ corresponds to stratifying the sample by W, calculating the mean in each of the two strata and then calculating a weighted average of these two means. Thus, the analysis model does not assume the population mean is the same in the two strata. If the imputation model does not include W, it assumes the population mean is the same in the two strata, with the result that the imputer is assuming more than the analyst, which is known to lead to overestimation of the variance of $\widehat{\boldsymbol{\theta}}_M$ when the extra assumption made by the imputer is correct (Meng, 1994). If the true value of the coefficients of $W\boldsymbol{Z}$ is zero, because the imputation model is correctly specified without the $W\boldsymbol{Z}$ terms, it is probably better not to include these terms and instead accept some overestimation of $\widehat{\boldsymbol{V}}_M$: imputation will be more efficient if they are set to zero rather than estimated.

Note that, because $\mathcal{R}(R) = 1$ only if $\boldsymbol{X}$ is complete, individuals with incomplete $\boldsymbol{X}$ are excluded, even if their Y and $\boldsymbol{Z}$ are complete. For this reason, it would not be appealing to use this method if the sample contained more than a few such individuals.

An alternative to IPW/MI is what we call "IPW/CC." Here Y is regressed on $\boldsymbol{Z}$ only in complete cases (those with $R_Y = 1$), again using weights W. This estimator is unbiased if

$$P\{R_Y = 1|Y, W, \boldsymbol{Z}, \mathcal{R}(R) = 1\} = P\{R_Y = 1|W, \boldsymbol{Z}, \mathcal{R}(R) = 1\}, \quad (6)$$

and the analysis model is correctly specified. If weights W are all equal, and $\boldsymbol{X} = \boldsymbol{Z}$ and the imputation and analysis models are the same, there is no benefit to IPW/MI over IPW/CC: it is more efficient to exclude individuals with missing Y (unless $M = \infty$, in which case exclusion and imputation are equivalent) (White and Carlin, 2010). However, there are two reasons for preferring IPW/MI to IPW/CC. These apply whether or not weights are equal. First, if (6) does not hold or if the analysis model is misspecified, the complete-case estimator may be inconsistent, whereas, as Theorem 2 states, IPW/MI gives consistent estimators if equation (5) holds (and assuming the imputation model is correctly specified). Equation (5) may be satisfied even if (6) is not, as (5) allows the probability that Y is observed to depend on a larger set of variables $\boldsymbol{X}$. Second, even if (6) holds and the analysis model is correctly specified, it may be more efficient to use all the available information (i.e., $\boldsymbol{X}$) to impute Y.

4. Simulation Study: Imputed Outcome

In this section, we explore IPW/MI for linear regression with imputed outcome. As in Section 3, the analysis model is fitted only to individuals with complete $\boldsymbol{X}$ and missing Y in these individuals are imputed.

Analysis of the sample must deal with two stages of missingness: stage 1 is the missingness in $\boldsymbol{X}$; stage 2, missingness in Y. At stage 1, one could either exclude individuals with incomplete $\boldsymbol{X}$ ($\mathcal{R}(R) = 0$) or impute missing $\boldsymbol{X}$. Similarly, each individual with missing Y not already excluded at stage 1 ($\mathcal{R}(R) = 1$) could either be excluded at stage 2 or have Y imputed. At each stage, if exclusion is used, one can either adjust for the exclusion using IPW or not adjust. Thus, there are three possibilities at each stage, giving $3 \times 3 = 9$ possible strategies in total. Denote a strategy by ST1/ST2, where ST1 and ST2 are each CC (exclude and do not weight), IPW (exclude and weight) or MI (impute). In IPW/MI, the focus of this article, individuals with missing $\boldsymbol{X}$ ($\mathcal{R}(R) = 0$) are excluded and weights used to adjust for this; individuals with complete $\boldsymbol{X}$ but missing Y ($\mathcal{R}(R) = 1$) have Y imputed. CC/CC uses only individuals with complete $\boldsymbol{X}$ and Y and there is no weighting. IPW/IPW uses the same individuals, but weights them by the inverse of their probability of being a complete case. In MI/MI all missing values are imputed. We also consider CC/IPW, CC/MI, and IPW/CC, but not MI/CC or MI/IPW, which combine the disadvantage of having to specify an imputation model for $\boldsymbol{X}$ with that of losing out on the potential efficiency gains of imputing Y.

The purpose of the following simulation is three-fold: to verify $\widehat{\boldsymbol{V}}$ is approximately unbiased for IPW/MI; to show IPW/MI can be more efficient than IPW/IPW; and to show MI/MI can yield biased parameter estimators when the stage 1 (for $\boldsymbol{X}$) or stage 2 (for Y given $\boldsymbol{X}$) imputation model is misspecified, and that IPW/MI remains approximately unbiased or at least less biased than MI/MI in these situations. The data-generating mechanism has been chosen to illustrate these points. It will now be described and then its features elucidated.

Data $\boldsymbol{X} = (X_1, X_2, X_3, X_4, X_5)$ and Y were generated for $N = 1000$ individuals. For each individual, X_1 was one with probability 0.5 and zero otherwise, X_2, X_3, and X_4 were independent and identically distributed $N(0, 1)$ and, finally, X_5 was sampled from $N(X_2 \times X_3, 1)$. Response Y was generated from

$$Y = -3 + X_1X_2 + X_1X_3 + 0.5X_2X_3 + X_4 + 0.5X_5 + \epsilon, \quad (7)$$

where $\epsilon \sim N(0, 1)$. X_1 was observed for all N individuals. With probability $0.8 - 0.6X_1$, (X_2, X_3, X_4, X_5) was observed; otherwise it was missing. If (X_2, X_3, X_4, X_5) was observed, Y was observed with probability $\{1 + \exp(-1.5 + 0.6X_2X_4)\}^{-1}$; otherwise Y was missing.

The analysis model was $Y = \theta_0 + \theta_2X_2 + \theta_3X_3 + \theta_{23}X_2X_3 + e$, where $E(e|X_2, X_3) = 0$. Therefore, $\boldsymbol{Z} = (1, X_2, X_3, X_2X_3)$. By integrating (7) with respect to X_1, X_4, and X_5, it can be shown that this analysis model is correctly specified and the true $\boldsymbol{\theta}$ is $(\theta_0, \theta_2, \theta_3, \theta_{23}) = (-3, 0.5, 0.5, 1)$.

This data-generating mechanism was chosen for three reasons. First, the X_1X_2 and X_1X_3 interactions in (7) mean the relation between Y and (X_2, X_3) is different in the two strata defined by X_1. Also, the probability that (X_2, X_3) is observed differs: in one stratum it is 0.2; in the other, 0.8. Thus, the relation between Y and (X_2, X_3) is different in individuals with complete $\boldsymbol{X}$ and incomplete $\boldsymbol{X}$. Failure to adjust for the missingness at stage 1, by weighting or imputation, will therefore lead to bias in θ_2 and θ_3. Therefore, CC/IPW, CC/MI, and CC/CC will be biased. Second, for individuals with observed (X_2, X_3, X_4, X_5) the probability Y is observed depends on X_4, which is not in the analysis model but is associated with Y. This causes the relation between Y and $\boldsymbol{X}$ described by the analysis model to be different in the set of complete cases from in the set with complete $\boldsymbol{X}$ but missing Y. In particular, because the probability of Y being missing depends on X_2X_4, the relation between Y and X_2 will be different in the two sets. Failure to adjust for the missingness at stage 2 will therefore lead to bias (specifically in θ_2). Therefore, IPW/CC, MI/CC, and CC/CC will be biased. Third, X_5 is included in the data-generating mechanism for Y to show that using MI at stage 1 can cause bias if the imputation model for $\boldsymbol{X}$ is misspecified (see results for MI*/MI below).

A total of 1000 datasets were generated and the seven methods applied to each. For each of θ_0, θ_2, θ_3, and θ_{23} and each method, the mean of the 1000 parameter estimates and of the 1000 estimated variances was calculated. The empirical SE was calculated as the standard deviation of the parameter estimates. Where a method involved imputation, 10 imputations were performed.

For MI/MI, the (correctly specified) imputation model at stage 1 was $(X_2, X_3, X_4) \sim N\{(\gamma_2, \gamma_3, \gamma_4), \Sigma_1\}$ and $X_5|X_2, X_3 \sim N(\gamma_5 + \gamma_6 X_2 + \gamma_7 X_3 + \gamma_8 X_2 X_3, \Sigma_2)$. Noninformative normal and inverse-Wishart priors were used, yielding normal and inverse-Wishart posteriors (Gelman et al., 2004, p. 88). For CC/MI, IPW/MI, and MI/MI, the (correctly specified) imputation model used at stage 2 was $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_{12} X_1 X_2 + \beta_{13} X_1 X_3 + \beta_{23} X_2 X_3 + \beta_{123} X_1 X_2 X_3 + \epsilon$.

For IPW/CC, IPW/IPW, and IPW/MI, weights were estimated by fitting the (correctly specified) missingness model for stage 1: $P(X_2, X_3, X_4$ and X_5 observed$) = \delta_0 + \delta_1 X_1$. Note that, because X_1 is binary, $W = (\delta_0 + \delta_1 X_1)^{-1} = \delta_0^{-1} - X_1 \delta_1 \{\delta_0(\delta_0 + \delta_1)\}^{-1}$ is a linear function of X_1. Hence, as the stage 2 imputation model includes $\boldsymbol{Z}$ and $X_1 \boldsymbol{Z}$, it implicitly includes $W\boldsymbol{Z}$. For CC/IPW and IPW/IPW, weights were estimated using the (correctly specified) model for stage 2: logit $P(Y$ observed $\mid \boldsymbol{X}$ observed$) = \delta_2 + \delta_3 X_2 + \delta_4 X_4 + \delta_5 X_2 X_4$. For IPW/IPW, the probability of being a complete case is the product of these two probabilities.

Table 1 shows mean parameter estimates, empirical SEs, and square roots of the mean estimated variances. It can be seen that IPW/MI yields approximately unbiased estimators of parameters and SEs, as expected from Theorem 2. As explained above, CC/IPW, CC/MI, CC/CC, and IPW/CC are biased for one or more parameters. IPW/IPW and MI/MI are both approximately unbiased. The former is less efficient than IPW/MI because the imputation model at stage 2 uses auxiliary information, i.e. covariates (notably X_4 and X_5) not included in the analysis model. The most efficient unbiased method is MI/MI, confirming that imputation is the best method when the imputation models are correct.

However, when the imputation model at stage 1 or stage 2 is misspecified, MI/MI may be biased. First, suppose that the imputation model at stage 1 is misspecified as $(X_2, X_3, X_4, X_5)^T \sim N\{(\gamma_2, \gamma_3, \gamma_4, \gamma_5)^T, \Sigma\}$. As X_2, X_3, X_4 and X_5 are uncorrelated (though not independent), Σ will be estimated as an approximately diagonal matrix. Therefore, for individuals with incomplete $\boldsymbol{X}$ the imputed values of X_5 will be approximately independent of X_2 and X_3; the relation between X_5 and the interaction of X_2 and X_3 ($E(X_5) = X_2 X_3$) is not present in the imputed data. The missing Y values of these individuals will then be imputed in such a way that the interaction between X_2 and X_3 is only 0.5, half what it should be. As half the individuals have incomplete $\boldsymbol{X}$, fitting the analysis model to the whole sample results in an estimate of θ_{23} of about 0.75. This is seen in Table 1 in the row MI*/MI.

Second, suppose the imputation model at stage 1 is correct but that at stage 2 is misspecified by leaving out the $\beta_{23} X_2 X_3$, $\beta_{123} X_1 X_2 X_3$, and $\beta_5 X_5$ terms. Missing Y values will now be imputed in such a way that there is no interaction between X_2 and X_3. As approximately 60% of Y values are missing, θ_{23} will be underestimated by about 60%. This result is shown in Table 1 in the row MI/MI*. The row IPW/MI* shows the result of IPW/MI with the same misspecified imputation model at stage 2. This method is considerably less biased than MI/MI*, because fewer Y values are being imputed. Therefore, the IPW element of IPW/MI provides some protection against misspecification of the imputation model.

5. Simulation Study: Imputed Covariate

In this section, we investigate the bias of $\widehat{\boldsymbol{V}}$ for IPW/MI in the case of linear regression with an imputed covariate. In the simulation study below, we find that the bias is small. This study also demonstrates again that IPW/MI can be more efficient than IPW/IPW, and that MI/MI can yield biased estimators when the imputation model for stage 1 is misspecified. Only brief details are presented here; full details can be found in the Web Appendix.

The (correctly specified) analysis model was $Y = \theta_0 + \theta_2 X_2 + \theta_3 X_3 + \theta_4 X_4 + \theta_{23} X_2 X_3 + e$, where $E(e \mid X_2, X_3, X_4) = 0$. Variables X_1 and Y were always observed; X_2 and X_3 were both observed or both missing. The probability they were observed depended on Y and X_1. If (X_2, X_3) was missing, so was X_4; otherwise the probability X_4 was observed depended on Y. The two stages of missingness are that stage 1 is missingness in (X_2, X_3) and stage 2 is missingness in X_4.

For MI/MI, the imputation model used at stage 1 (to impute X_2 and X_3) falsely assumed that (Y, X_2, X_3) was trivariate normal. Although misspecified, this imputation model might easily be used in practice. As the stage 1 imputation model is misspecified, we call this method MI*/MI. For IPW/MI and MI*/MI, the imputation model used at stage 2 (to impute X_4) was correctly specified in terms of X_1, X_2, X_3, Y, and certain interactions. The covariates (X_1 and Y) that determine the weights are included in this model. IPW/MI* and MI*/MI* used a stage 2 imputation model that was misspecified because interaction terms were omitted.

Table 2 shows the results. IPW/IPW and IPW/MI are approximately unbiased, and SE estimators for IPW/MI are approximately unbiased. SEs for IPW/MI are smaller than for IPW/IPW: it is more efficient to impute missing X_4 for individuals with otherwise complete data than to exclude them.

MI*/MI gives biased estimation, because the imputation model at stage 1 is misspecified. Misspecification also of the imputation model at stage 2 (MI*/MI*) adds to the bias, especially in θ_4. Bias also occurs when IPW is used at stage 1 instead of MI (IPW/MI*), but is smaller than that of MI*/MI*, and indeed of MI*/MI.

Theorems 1 and 2 of Robins and Wang (2000) enable the asymptotic ($N \to \infty$) percentage bias in $\widehat{\boldsymbol{V}}_M$ to be calculated when $M = \infty$ (see Web Appendix). The asymptotic percentage bias in $\widehat{\boldsymbol{V}}_M$ was 3.7% for θ_4 and less than 1% for θ_0, θ_2, θ_3, and θ_{23}, which is in line with the finding above that $\widehat{\boldsymbol{V}}_M$ was approximately unbiased for finite N and M.

The results above were obtained using the true weights. In practice, weights would usually be estimated. Row IPWe/MI in Table 2 shows the results when weights are estimated. The variance estimators are approximately unbiased. Note that for IPWe/MI, $\widehat{\boldsymbol{V}}$ was replaced by a sandwich estimator that accounts for uncertainty in the weights (Robins et al., 1994). When $\widehat{\boldsymbol{V}}$ was instead used, the variance for θ_0 was overestimated.

6. Application

The NCDS consists of 17,638 individuals born in Britain during one week in 1958 (Power and Elliott, 2006). 920 immigrants added later are not considered here. Data were collected at birth and at ages 7, 11, 16, 23, 33, and 45. A total

Table 1

Mean parameter estimate ("mean"), square root of mean estimated variance ("aSE"), and empirical SE ("eSE") for four parameters and 10 analysis methods. The true value of $\boldsymbol{\theta}$ is $(\theta_0, \theta_2, \theta_3, \theta_{23}) = (-3, 0.5, 0.5, 1)$.

	θ_0			θ_2			θ_3			θ_{23}		
Method	Mean	aSE	eSE	Mean	aSE	eSE	Mean	aSE	eSE	Mean	aSE	eSE
True	−3.000			.500			.500			1.000		
CC/CC	−2.995	.080	.079	.090	.081	.087	.200	.080	.086	1.005	.082	.091
CC/IPW	−2.993	.082	.079	.199	.092	.091	.200	.086	.089	1.004	.094	.100
CC/MI	−2.994	.075	.076	.202	.081	.081	.201	.079	.083	1.004	.084	.086
IPW/CC	−2.993	.102	.101	.382	.110	.112	.495	.109	.114	1.008	.114	.119
IPW/IPW	−2.990	.106	.104	.489	.120	.124	.494	.112	.117	1.006	.121	.132
IPW/MI	−2.992	.097	.096	.498	.105	.105	.497	.104	.107	1.006	.110	.113
MI/MI	−3.000	.089	.081	.503	.092	.087	.497	.090	.088	1.006	.092	.082
MI*/MI	−2.998	.092	.085	.498	.095	.093	.496	.094	.094	.749	.100	.083
MI/MI*	−2.999	.108	.101	.100	.088	.054	.099	.088	.051	.391	.091	.055
IPW/MI*	−2.998	.107	.100	.492	.119	.122	.495	.117	.115	.776	.131	.127

Table 2

Mean parameter estimate (mean), square root of mean estimated variance (aSE), and empirical SE (eSE) for five parameters and 10 analysis methods. Results for θ_2 are omitted because, apart from Monte Carlo error, they are the same as for θ_3. The true value of $\boldsymbol{\theta}$ is $(\theta_0, \theta_2, \theta_3, \theta_4, \theta_{23}) = (0, 0.5, 0.5, 0.5, 1)$.

	θ_0			θ_3			θ_4			θ_{23}		
Method	Mean	aSE	eSE	Mean	aSE	eSE	Mean	aSE	eSE	Mean	aSE	eSE
True	.000			.500			.500			1.000		
CC/CC	.238	.060	.056	.196	.061	.065	.183	.060	.064	.992	.064	.077
IPW/IPW	.020	.095	.102	.485	.103	.113	.479	.108	.124	.990	.108	.119
IPW/MI	.002	.075	.075	.495	.084	.084	.490	.092	.089	1.001	.089	.088
MI*/MI	−.086	.051	.061	.663	.100	.129	.372	.071	.072	.976	.079	.117
MI*/MI*	−.087	.051	.060	.674	.100	.126	.337	.077	.081	.970	.080	.112
IPW/MI*	−.003	.078	.076	.504	.086	.091	.427	.096	.089	.978	.092	.095
IPWe/MI	.003	.061	.060	.497	.081	.083	.491	.089	.087	1.001	.088	.089

of 16,334 nonimmigrants were still alive and free from type 1 diabetes at age 45 and of these, 8953 (55%) participated in a biomedical survey.

Thomas, Hypponen, and Power (2007) investigated the effect of characteristics measured at birth and adult adiposity (body mass index [BMI] and waist size at 45) on glucose metabolism at age 45. Subjects were classified as having high blood glucose if their glycosylated hemoglobin (A1C) was greater than 6% or they had type 2 diabetes. Immigrants and individuals with type 1 diabetes were excluded. Data on blood glucose, BMI and waist size at 45 were available for 7518 of the 8953 participants. Of these, 1845 (25%) had incomplete data on the factors measured at birth. Thomas et al., using the ice command in STATA (Royston, 2005), performed MI by chained equations (Van Buuren, 2007) on the 7518 subjects, producing 10 complete datasets. These 7518 were then analyzed as though representative of all 16,334 nonimmigrants alive and free from type 1 diabetes at age 45. Thomas et al. concluded that the factors measured at birth were related to blood glucose at 45 and that, moreover, some of these effects were largely mediated through adult adiposity.

We repeated this analysis but used IPW to allow the relation between glucose and the predictors to differ in the 7518 subjects with complete age 45 data from the other 8816 cohort members. Here stage 1 missingness refers to the age 45 data and stage 2 refers to the data measured at or before birth. Thomas et al. used a CC/MI analysis (i.e., used complete cases at stage 1 and MI at stage 2), whereas we use IPW/MI.

In the missingness model for stage 1, i.e., for the probability that at least one of glucose, BMI and waist size is missing, we used the potential predictors of missingness recorded at birth or age 7 identified by Atherton et al. (2008) and listed in their Table 3. We also used gestational age (< 38 versus ≥ 38 weeks) and a set of variables recorded at age 11: math and reading scores (normal/low), internalizing and externalizing problems (normal/intermediate/problem), and verbal and nonverbal scores (normal/low). All predictors were categorical, and most binary.

Not everyone attended the age 7 and age 11 visits, and even those who did had some missing values. Therefore, some predictors of missingness at stage 1 were themselves missing. To deal with visit missingness, we partitioned the sample into four strata according to which of the age 7 and age 11 visits were attended. A different logistic regression was fitted to each stratum, using only predictors from the visits attended by individuals in that stratum. Missing values in these predictors were dealt with by introducing missing indicator variables. The missing indicator method can cause bias when used for variables in an analysis model (Jones, 1996). Although we are

Table 3

LOR and SEs for predictors of high blood glucose. Binary predictors are gestational age < 38 weeks, preeclampsia, smoking during pregnancy, prepregnancy BMI $\geq 25 Kg/m^2$, and manual socioeconomic position (SEP) at birth. Ordinal and continuous predictors are birth weight for gestational age (tertile), BMI at age 45 (Kg/m^2), and waist circumference at age 45 (cm). Adjustment was also made for sex and family history of diabetes.

	CC/MI		IPW/MI		MI/MI	
	LOR	SE	LOR	SE	LOR	SE
Short gestation	0.46	0.22	0.48	0.23	0.44	0.20
Preeclampsia	0.46	0.27	0.55	0.27	0.47	0.25
Mother overweight	0.29	0.15	0.36	0.16	0.18	0.12
Smoke in pregnancy	0.02	0.14	0.04	0.14	0.04	0.14
Manual SEP	0.37	0.17	0.44	0.18	0.39	0.17
Birth weight	−0.31	0.09	−0.31	0.09	−0.32	0.09
BMI age 45	0.04	0.02	0.02	0.02	0.03	0.02
Waist size age 45	0.07	0.01	0.07	0.01	0.07	0.01

using it to calculate weights, not in the analysis model, this method is imperfect and we do not recommend it for general use. Therefore, we also calculated a second set of weights by multiply imputing missing predictors of missingness. The results obtained using this second set of weights were very similar to those (reported below) obtained using missing indicators.

The mean weight was 2.5; 5th and 95th percentiles were 1.6 and 5.2; the maximum was 23.1. As found by Atherton et al. (2008), disadvantaged individuals were more likely to be missing at stage 1. In the stratum who attend both age 7 and 11 visits, the following variables were significant at the 5% level: breastfed $<$1 month; mother leaving school at or before statutory age; short stature, overweight, internalizing, and externalizing problems at age 7; internalizing and externalizing problems, low math, low reading, and low nonverbal scores at age 11.

For stage 2 we used the same imputation model as Thomas et al., except that we included the weights. Following guidelines of White, Royston, and Wood (2010), 25 imputations were used. This MI model used only the variables in the analysis model and the weights. We also tried adding variables used as predictors in the missingness model to the imputation model, but this made very little difference to the results below.

Table 3 shows the estimated log odds ratios (LOR) and SEs. Due to the stochastic nature of MI and the inclusion of weights in the imputation model, the results for CC/MI are slightly different (maximum difference 0.03) from those reported by Thomas et al. (2007). As can be seen, using IPW at stage 1 (IPW/MI) does not substantially change the results. The biggest differences are that ORs for preeclampsia, mother overweight, and manual class have risen slightly, and the first two have changed from being almost significant to just significant. SEs are also slightly larger.

We investigated why these ORs increased slightly when weighting was used. The missingness model indicated that disadvantaged individuals were more likely to be missing at stage 1. Therefore, using IPW gives more weight to disadvantaged individuals. We partitioned the stratum who attended both age 7 and age 11 visits into two groups, advantaged and disadvantaged, using the following rule: individuals with at least three of the following indicators of disadvantage were classified as disadvantaged: breastfed < 1 month; mother leaving school early; short stature, overweight, internalizing, and externalizing problems at age 7; and internalizing and externalizing problems, and poor math, reading, and nonverbal scores at age 11. Using this rule, the disadvantaged group contained 29% of individuals. The other 71% were classified as advantaged. The analysis model was fitted to the two groups separately. The LORs for preeclampsia, mother overweight, and social class were 0.59, 0.70, and 0.33, respectively, in the disadvantaged group, and −0.04, −0.04, and 0.41 in the advantaged group. Therefore, the observed relation between glucose and preeclampsia/overweight is stronger in the disadvantaged individuals. It seems likely therefore that the reason why ORs for preeclampsia and overweight in the whole cohort are greater when IPW is used (IPW/MI versus CC/MI) is that IPW gives more weight to the disadvantaged group. The relation between manual class and glucose, however, is slightly weaker in the disadvantaged group, leaving its increased OR unexplained.

Assuming then that the probability that glucose, BMI and waist size at 45 years are complete does not depend on variables in the analysis model given available predictors of missingness, the associations found by Thomas et al. in the sample of 7518 individuals do generalize to the population of nonimmigrants still alive and free from type 1 diabetes at age 45.

Finally, we used MI/MI, i.e., imputed all missing values for all 16,334 individuals. Included in the imputation were the variables in the analysis model and the predictors in the missingness model of IPW/MI. A total of 100 imputed datasets were created. Table 3 shows the results. They do not differ substantially from those of IPW/MI. Some SEs are slightly smaller. The small increases in the ORs of preeclampsia and mother overweight seen in IPW/MI relative to CC/MI are not replicated. In fact, the OR of overweight is lower in MI/MI than in CC/MI.

To investigate why, we partitioned the 12,501 individuals who attended both age 7 and age 11 visits into four groups, using the same rule for disadvantage as before: disadvantaged with observed glucose; disadvantaged with imputed glucose; advantaged with observed glucose; and advantaged with imputed glucose. The analysis model was fitted to each group separately. It was found that, whereas the relation between blood glucose and its predictors differed considerably between the advantaged and disadvantaged groups in the set of individuals whose glucose was observed, this difference was not seen in those with imputed glucose. In particular, the LORs for overweight were 0.57 and -0.17 in the disadvantaged and advantaged groups with observed glucose, respectively, but were 0.15 and 0.18 for those with imputed glucose. Interaction terms are needed in the imputation model, e.g., imputation could be done separately in the two groups. Careful assessment of the imputation model might have revealed this, but such assessment might not always be made.

7. Discussion

Robins and Wang (2000) derive a general formula for the asymptotic variance of an MI estimator based on a complete-data estimator solving a set of estimating equations. This formula applies when improper imputation and a parametric imputation model are used. IPW/MI could be carried out in this way and the Robins and Wang (2000) variance formula used. The formula is, however, complicated and has not been implemented in standard software. Using proper imputation with Rubin's rules is appealing because it is simpler and can be used with nonparametric imputation procedures. Robins and Wang (2000) also give a formula for the asymptotic bias of the Rubin's rules variance estimator when $M = \infty$. We used this to show that, in the case of linear regression with MI of a missing outcome, the Rubin's rules variance estimator for IPW/MI is consistent when $M = \infty$. We also used it in the setting described in Section 5, where a missing covariate is imputed. The expression derived for the asymptotic bias in the Rubin's rules variance estimator for IPW/MI was complicated and did not reduce to zero. However, both the asymptotic and finite-sample biases were found to be small in this study. In the Web Appendix, we describe two simulation studies of logistic regression, one with an imputed outcome and one with an imputed covariate. In both, the Rubin's rules variance estimator was approximately unbiased. Schafer (2003) comments that "although we may find it difficult to prove good performance for [MI using a nonmaximum likelihood estimator], that does not imply that good performance will not be seen in practice. Experience suggests that Bayesian MI does interact well with a variety of semi- and nonparametric estimation procedures."

If the weights are just sampling weights, they will be known, but if they are used to account for missing data, they will need to be estimated. A limitation of our proof in Section 4 is that the complete-data variance estimator assumes that weights are known and ignores any estimation uncertainty about them. This uncertainty is commonly ignored, thus overestimating the variance (Robins et al., 1994), as we saw in Section 5. If software allows, we recommend using a sandwich estimator that accounts for the uncertainty in the weights (Robins et al., 1994).

Some researchers may prefer to use straightforward MI (what we called MI/MI). Provided that the imputation models are correctly specified, this will be more efficient than IPW/MI. However, our (admittedly contrived) simulations and (not contrived) real data example have shown that those who prefer IPW/MI have some justification for their caution. A possible use for IPW/MI is as a check, or diagnostic, for MI/MI. If the results of IPW/MI and MI/MI are very different, further exploration would be warranted, possibly leading to refinement of the imputation model. We have not considered the effect of misspecified missingness models. Such misspecification would typically cause bias, just like misspecification of the imputation model in MI/MI. However, the fit of the missingness model, which is a model for a univariate response, is easier to assess, and more able to be assessed (Vansteelandt, Carpenter, and Kenward, 2010), than that of a complex multivariate imputation model. Furthermore, IPW/MI is needed when sampling weights are used, even if all missing values are imputed.

IPW/MI will be most appealing when the model for the weights is relatively simple compared with the imputation model. This will not always be so. Also, a limitation of all IPW methods is their difficulty in handling nonmonotone missingness in the predictors in the missingness model. Robins and Gill (1997) propose a procedure for handling such missingness, but this is complicated to use and limited in practice to a small number of missing predictors.

Another alternative to IPW/MI is IPW/IPW. This is simpler, but has the disadvantage that an individual is excluded from an analysis even if he/she is missing just one variable. Furthermore, if multiple analyses are being performed with different variables, either a different set of weights is needed for each analysis (because an individual who is complete for one analysis may be incomplete for another) or a single set of weights is calculated but only for individuals who are complete cases for all the analyses (Goldstein, 2009). IPW/MI, on the other hand, would allow a single set of weights to be used, as imputation could ensure that the set of complete cases were the same for each analysis.

8. Supplementary Materials

The Web Appendix referenced in Sections 2, 3, 5, and 7 is available under the Paper Information link at the *Biometrics* website `http://www.biometrics.tibs.org/`.

Acknowledgements

We thank Chris Power for valuable discussion and assistance in obtaining the NCDS data, Claudia Thomas for preparing the variables we used, and two anonymous reviewers. The Centre for Longitudinal Studies provided the official NCDS data. SS and IW were funded by MRC grants MC_US_A030_0014 and MC_US_A030_0015 and LL by an MRC Career Development Award.

References

Atherton, K., Fuller, E., Shepherd, P., Strachan, D. P., and Power, C. (2008). Loss and representativeness in a biomedical survey at age 45 years: 1958 British Birth Cohort. *Journal of Epidemiology and Community Health* **62,** 216–223.

Caldwell, T. M., Rodgers, B., Clark, C., Jefferis, B. J. M. H., Stansfeld, S. A., and Power, C. (2008). Lifecourse socioeconomic predictors of midlife drinking patterns, problems and abstention: Findings from the 1958 British Birth Cohort study. *Drug and Alcohol Dependence* **95,** 269–278.

Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (2004). *Bayesian Data Analysis*. London: Chapman and Hall/CRC.

Goldstein, H. (2009). Handling attrition and non-response in longitudinal data. *Longitudinal and Life Course Studies* **1,** 63–72.

Höfler, M., Pfister, H., Lieb, R., and Wittchen, H. (2005). The use of weights to account for non-response and drop-out. *Social Psychiatry and Psychiatric Epidemiology* **40,** 291–299.

Jones, M. P. (1996). Indicator and stratification methods for missing explanatory variables in multiple linear regression. *Journal of the American Statistical Association* **91,** 222–230.

Kim, J. K., Brick, J. M., Fuller, W. A., and Kalton, G. (2006). On the bias of the multiple-imputation variance estimator in survey sampling. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **68,** 509–521.

Little, R. J. A. and Rubin, D. B. (2002). *Statistical Analysis with Missing Data*. New Jersey, NJ: Wiley.

Meng, X.-L. (1994). Multiple-imputation inferences with uncongenial sources of input. *Statistical Science* **9,** 538–573.

Nielsen, S. F. (2003). Proper and improper multiple imputation. *International Statistical Review* **71,** 593–627.

Power, C. and Elliott, J. (2006). Cohort profile: 1958 British Birth Cohort (National Child Development Study). *International Journal of Epidemiology* **35,** 34–41.

Priebe, S., Fakhoury, W., White, I., Watts, J., Bebbington, P., Billings, J., Burns, T., Johnson, S., Muijen, M., Ryrie, I., Wright, C., and P.L.A.O.S. Group (2004). Characteristics of teams, staff and patients: Associations with outcomes of patients in assertive outreach. *British Journal of Psychiatry* **185,** 306–311.

Robins, J. M. and Gill, R. D. (1997). Non-response models for the analysis of non-monotone ignorable missing data. *Statistics in Medicine* **16,** 39–56.

Robins, J. and Wang, N. (2000). Inference for imputation estimators. *Biometrika* **87,** 113–24.

Robins, J., Rotnitzky, A., and Zhao, L. (1994). Estimation of regression coefficients when some regressors are not always observed. *Journal of the American Statistical Association* **89,** 846–866.

Royston, J. (2005). Multiple imputation of missing values: Update of ice. *Stata Journal* **5,** 527–536.

Rubin, D. B. (1987). *Multiple Imputation for Nonresponse in Surveys*. New York, NJ: Wiley.

Schafer, J. L. (2003). Multiple imputation in multivariate problems when the imputation and analysis models differ. *Statistica Neerlandica* **57,** 19–35.

Schenker, N. and Welsh, A. H. (1988). Asymptotic results for multiple imputation. *Annals of Statistics* **16,** 1550–1566.

Stansfeld, S. A., Clark, C., Caldwell, T. M., Rodgers, B., and Power, C. (2008a). Psychosocial work characteristics and anxiety and depressive disorders in midlife: The effects of prior psychological distress. *Occupational and Environmental Medicine* **65,** 634–642.

Stansfeld, S. A., Clark, C., Rodgers, B., Caldwell, T. M., and Power, C. (2008b). Childhood and adulthood socio-economic position and midlife depressive and anxiety disorders. *Drug and Alcohol Dependence* **95,** 269–278.

Thomas, C., Hypponen, E., and Power, C. (2007). Prenatal exposures and glucose metabolism in adulthood. *Diabetes Care* **30,** 918–924.

Van Buuren, S. (2007). Multiple imputation of discrete and continuous data by fully conditional specification. *Statistical Methods in Medical Research* **16,** 219–242.

Vansteelandt, S., Carpenter, J., and Kenward, M. G. (2010). Analysis of incomplete data using inverse probability weighting and doubly robust estimators. *Methodology* **6,** 37–48.

Wang, N. and Robins, J. M. (1998). Large-sample theory for parametric multiple imputation procedures. *Biometrika* **85,** 935–948.

White, I. R. and Carlin, J. B. (2010). Bias and efficiency of multiple imputation compared with complete-case analysis for missing covariate values. *Statistics in Medicine* **29,** 2920–2931.

White, I. R., Royston, P., and Wood, A. M. (2010). Multiple imputation using chained equations: Issues and guidance for practice. *Statistics in Medicine* **30,** 377–399.

Received October 2009. *Revised June* 2011.
Accepted June 2011.