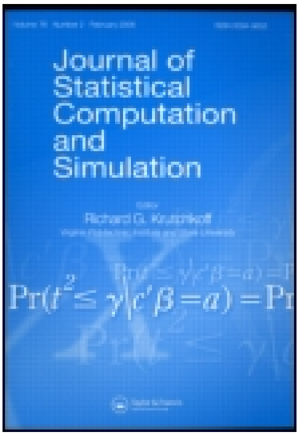

Journal of Statistical Computation and Simulation
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/gscs20

Combining the complete-data and nonresponse models for drawing imputations under MAR

S. Jolani [a] , S. van Buuren [a b] & L. E. Frank [a]
[a] Department of Methodology and Statistics, Faculty of Social and Behavioral Sciences , Utrecht University , Utrecht , The Netherlands
[b] Department of Statistics, TNO Quality of Life , Leiden , The Netherlands
Published online: 07 Dec 2011.

To cite this article: S. Jolani , S. van Buuren & L. E. Frank (2013): Combining the complete-data and nonresponse models for drawing imputations under MAR, Journal of Statistical Computation and Simulation, 83:5, 868-879

To link to this article: http://dx.doi.org/10.1080/00949655.2011.639773

Journal of Statistical Computation and Simulation, 2013
Vol. 83, No. 5, 868–879, http://dx.doi.org/10.1080/00949655.2011.639773

Combining the complete-data and nonresponse models for drawing imputations under MAR

S. Jolani[a]*, S. van Buuren[a,b] and L.E. Frank[a]

[a]*Department of Methodology and Statistics, Faculty of Social and Behavioral Sciences, Utrecht University, Utrecht, The Netherlands;* [b]*Department of Statistics, TNO Quality of Life, Leiden, The Netherlands*

(*Received 7 July 2011; final version received 8 November 2011*)

In multiple imputation (MI), the resulting estimates are consistent if the imputation model is correct. To specify the imputation model, it is recommended to combine two sets of variables: those that are related to the incomplete variable and those that are related to the missingness mechanism. Several possibilities exist, but it is not clear how they perform in practice. The method that simply groups all variables together into the imputation model and four other methods that are based on the propensity scores are presented. Two of them are new and have not been used in the context of MI. The performance of the methods is investigated by a simulation study under different missing at random mechanisms for different types of variables. We conclude that all methods, except for one method based on the propensity scores, perform well. It turns out that as long as the relevant variables are taken into the imputation model, the form of the imputation model has only a minor effect in the quality of the imputations.

Keywords: dual modelling; missingness mechanism; misspecification; multiple imputation; propensity score

AMS Subject classifications: F1.1; F4.3

1. Introduction

Multiple imputation (MI) is an important and influential approach in the analysis of missing data. Although proposed in the context of nonresponse in sample surveys at first, the technique is quite general and can be readily used in other settings as well. The merits of MI and recent software developments have been discussed elsewhere [1–3]. Generally, MI works as follows: every missing datum is replaced by two or more imputed values in order to reflect uncertainty about which value to impute and then each completed data set is analysed separately by standard statistical tools just as if the imputed data were real. Estimates of parameters are combined by using Rubin's rule [4, p. 76] to make the final inferences about the data.

The validity of MI depends on the imputed values. If an inappropriate model for the imputation process is used, it will result in biased and inconsistent conclusions [5]. A full understanding of the methodology to obtain imputed values is important to obtain unbiased estimates with correct confidence intervals. Failure to consider all relevant aspects of creating imputation models can

*Corresponding author. Email: s.jolani@uu.nl

impact the validity of inferences. For example, if our scientific interest focuses on the correlation between two variables, then both variables must be present in the imputation process even if only one of them has missing values. If we mistakenly remove the other variable from the imputation model, then inferences derived from MI will be biased. In general, the imputation model should not impose unnecessary restrictions on the variables that will be the subject of the later analysis. An imputation model should also preserve the relations among the variables in the post-imputation analysis. As a general guideline, imputation methods should be general enough to encompass the intended analyses.

To produce high-quality imputed values for a particular incomplete variable, the imputation model should include variables that are (i) potentially related to the incomplete variable and (ii) potentially related to the missingness of the incomplete variable. It should be noted that a particular variable in this case may have different roles: it is part of the complete-data model, the nonresponse model or both.

There are many ways to combine both types of variables in the complete-data model and the nonresponse model, yet it is not clear how these models should be represented in the imputation model. The most commonly used technique is to simply group variables together in both the complete-data and the nonresponse models as predictors [6–8]. Propensity score methods are another way to incorporate the nonresponse model into the complete-data model for drawing imputed values [9,10]. It is also possible to apply a dual modelling strategy [10,11]. It can be done by adding an appropriate function of the propensity scores into the imputation model. We introduce two new methods that have not been used in MI.

There appears to be no consensus among statisticians about which method is best. A particular variable may have different roles and it is not well known how to combine variables that have different roles. As far as we know, no research has been done yet to evaluate methods for combining variables in the complete-data and nonresponse models.

In this paper, we compare the performance of several strategies to handle different types of variables by a simulation study. The methods are evaluated under different conditions of missing at random (MAR) and also for different kinds of variables (i.e. continuous, dichotomous and polytomous). The goal of the paper is to help researchers to make informed choices in their applications.

This article is organized as follows. In the next section, we provide the notation used in this paper. In Section 3, we define several methods to create imputation models for drawing imputed values. Evaluation of the imputation methods is studied by simulations in Section 4. The discussion is given in the last section.

2. Notation

Suppose we have a study that collects observations for n subjects on p different variables $\mathbf{Z} = (Z_1, \ldots, Z_p)'$. In practice, for some subjects, some of these observations are missing. The observed and missing part of each subject is defined as $\mathbf{Z}_{\text{obs}}$ and $\mathbf{Z}_{\text{mis}}$, respectively. Assume our primary interest is to estimate (population) parameters θ, such as means, correlations, regression coefficients. For this, we have a substantive model, which is called the complete-data model. Denote $\mathbf{X}$ as all variables in this model where some of them are outcomes and the others are explanatory variables. In order to obtain valid inferences under the assumed missing data mechanism, we specify a (probabilistic) structure for the missingness. An indicator variable $\mathbf{R} = (R_1, \ldots, R_p)'$ is defined to show the occurrence of the missing data, where $R_j = 1$ ($j = 1, \ldots, p$) indicates that the corresponding value of $\mathbf{Z}$ is observed. $\mathbf{R}$ is subject to a probability distribution $P(\mathbf{R}|\mathbf{U})$, where $\mathbf{U}$ indicates variables that are known to have influenced the occurrence of missing data. Thus, the nonresponse model can be expressed as $\mathbf{R} = g(\mathbf{U}; \psi)$, where ψ represents unknown parameters in the nonresponse model. Variables in the complete-data model and the nonresponse model may also share some common variables, that is, $\mathbf{X}$ and $\mathbf{U}$ are possibly overlapping subsets of $\mathbf{Z}$.

3. Imputation methods

The specification of the imputation model is the most complex step in MI. When the missingness mechanism is MAR, which means the probability of missingness depends only on the observed part of data [12], the main task is to determine the posterior predictive distribution of $\mathbf{Z}_{\text{mis}}$ given $\mathbf{Z}_{\text{obs}}$. In this section, we present several methods for specifying it. Consider a univariate model where missing values belong only to the outcome variable. Let us define the complete-data model $Y_i = \mathbf{X}_i\beta_x + \epsilon_i$ and the corresponding nonresponse model $R_i = \mathbf{U}_i\beta_u + \tau_i$ for $i = 1, \ldots, n$. Table 1 gives a summary of the methods. Each method will be discussed in detail in the following sections.

3.1. *Additional predictors*

The variables of the nonresponse model are used as additional predictors in the imputation model [6]. Assume $\mathbf{X}_{\text{obs}}$ and $\mathbf{U}_{\text{obs}}$ denote observed parts of variables in the incomplete-data model and the nonresponse model, respectively. Imputed values from the posterior predictive distribution $P(\mathbf{Z}_{\text{mis}}|\mathbf{X}_{\text{obs}}, \mathbf{U}_{\text{obs}})$ are drawn as outlined in [4, Chapter 5]. Assume a prior distribution for the parameters θ and calculate the posterior distribution of θ given the observed data, $P(\theta|\mathbf{X}_{\text{obs}}, \mathbf{U}_{\text{obs}})$. Then, a value of θ^* is drawn from its posterior distribution, and values of $\mathbf{Z}^*_{\text{mis}}$ are drawn from the conditional predictive distribution $P(\mathbf{Z}_{\text{mis}}|\mathbf{X}_{\text{obs}}, \mathbf{U}_{\text{obs}}, \theta = \theta^*)$. Repeating these steps separately m times produces m completed data sets.

3.2. *Propensity covariate*

Using the propensity scores [13] offers researchers a particularly desirable way for adjusting the effects of variables in the nonresponse model into the imputation process. The basic strategy is to find a single number summary of the covariates in the nonresponse model that will remove bias, if any, from the final estimates in the post-imputation analysis. Propensity scores are defined as the probability of being observed as a function of the covariates in the nonresponse model. These scores are unknown and can be estimated from the sample.

A common way to estimate propensity scores is to use a logistic regression model. For the jth incomplete variable, it assumes $\log[\pi_{ij}/(1-\pi_{ij})] = \mathbf{U}_{ij}\psi_j$, where $\mathbf{U}_{ij}$ is a vector of covariates in the nonresponse model for ith individual and the jth incomplete variable, $\pi_{ij} = P(R_{ij} = 1|\mathbf{U}_{ij})$, the probability of being observed given covariates $\mathbf{U}_{ij}$, and ψ_j is a vector of parameters that are

Table 1. Imputation methods for a univariate incomplete variable y.

Abbreviation	Method	Model
AP	Additional predictors	$Y = \alpha + \mathbf{X}_{\text{obs}}\beta_x + \mathbf{U}_{\text{obs}}\beta_u + \varepsilon$
PC	Propensity covariate	$Y = \alpha + \mathbf{X}_{\text{obs}}\beta_x + \hat{\pi}\beta_\pi + \varepsilon$
PS	Propensity stratification	$Y = \alpha_s + \mathbf{X}_{s,\text{obs}}\beta_{x_s} + \varepsilon$
IPC	Inverse propensity covariate	$Y = \alpha + \mathbf{X}_{\text{obs}}\beta_x + \hat{\pi}^{-1}\beta_{\pi^{-1}} + \varepsilon$
PDV	Propensity dummy variable	$Y = \alpha + \mathbf{X}_{\text{obs}}\beta_x + \hat{\pi}_s\beta_{\pi_s} + \varepsilon$

usually estimated by maximum likelihood. Solving for π_{ij} gives

$$\pi_{ij} = \frac{\exp(\mathbf{U}_{ij}\psi_j)}{1 + \exp(\mathbf{U}_{ij}\psi_j)}, \quad (1)$$

for $i = 1, \ldots, n$. Estimated propensity scores are obtained by inserting the estimated values for the parameters in Equation (1).

Assuming that all relevant covariates have been taken into account, the propensity scores can be directly added into the imputation model as an extra predictor. An obvious advantage is the ability to summarize multi-dimensional covariates into a one-dimensional covariate. With dozens or hundreds of variables, it is often difficult to use all variables in an imputation model. To impute the missing values of a particular variable Z_j, the imputation model is defined by $P(Z_{j,\mathrm{mis}}|\mathbf{X}_{\mathrm{obs}}, \hat{\pi}_j)$, where $\hat{\pi}_j$ is a vector of the estimated propensity scores for the jth incomplete variable. A similar principle will be used for other incomplete variables.

3.3. *Propensity stratification*

Some researchers prefer to coarsen the propensities into a few classes in which missing and observed individuals have homogenous propensity scores [10]. If we are able to create homogenous classes, then, in theory, missing and observed individuals within each class have an identical distribution for the incomplete variable. This means that dividing the population into groups of constant propensity may remove bias due to the nonresponse model. Lavori *et al.* [9] suggested that one categorizes the estimated propensity scores into five categories and then apply a nonparametric imputation method, the approximate Bayesian bootstrap [4, p.44], within each category. Here, we use a slightly different approach. A separate imputation model is defined within each category. Assume $\mathbf{Z}_{s,\mathrm{mis}}$ represents the missing part of the data in the category s. The imputation model is then presented as $P_s(\mathbf{Z}_{s,\mathrm{mis}}|\mathbf{X}_{s,\mathrm{obs}})$, where $\mathbf{X}_{s,\mathrm{obs}}$ denotes the observed part of variables in the complete-data model. Drawing imputations and replacing them within each category provide completed data sets that are used in the subsequent analysis.

3.4. *Inverse propensity covariate*

During the last two decades, there has been growing interest in dual modelling strategies [14–17]. Estimation based on dual modelling strategies has an interesting property known as double robustness, which will be explained later in this section. The idea is that the estimated parameter remains consistent if either a model for the missingness mechanism or a model for the distribution of the complete-data (but not necessarily both) has been specified correctly. Several authors considered dual modelling strategies for parameter estimation in missing data analysis [10,11]. Carpenter *et al.* [18] compared MI and doubly robustness for parameter estimation in the presence of missing data.

We propose to apply a dual modelling strategy in the imputation procedure in the following way. Reconsider the complete-data model $Y_i = \mathbf{X}_i\beta + \epsilon_i$, where

$$E(Y_i|\mathbf{X}_i) = \mathbf{X}_i\beta, \quad (2)$$

with $E(\epsilon_i) = 0$. An obvious candidate to impute missing values based on Equation (2) is $\mathbf{X}_i\hat{\beta}$ (plus some random noise), where $\hat{\beta}$ is the estimator of the coefficients from the regression of Y_i on $\mathbf{X}_i$. If the regression model is specified correctly, then the mean of the estimated residuals in the population will be zero. In most cases, however, the regression model is a rough approximation to the true regression model. The implication of this misspecification can be serious if $P(Y_i|\mathbf{X}_i, R_i = 1)$ is different from $P(Y_i|\mathbf{X}_i, R_i = 0)$.

Scharfstein *et al.* [14, pp. 1140–1141] showed that to obtain a doubly robust estimator for the population mean, it suffices to add the inverse of the propensity scores for each individual, π_i^{-1}, to the regression model. Here, we suggest to include this term as an extra predictor into the imputation model for the incomplete variable. The imputation model for our example can be written as

$$E(Y_i|\mathbf{X}_i, \mathbf{U}_i) = \mathbf{X}_i\beta + \gamma\hat{\pi}_i^{-1},$$

where $\hat{\pi}_i$ are the estimated propensity scores obtained by Equation (1) and $(\beta, \gamma)'$ is a vector of coefficients which can be estimated by several approaches such as least squares.

If the complete-data model correctly describes $E(Y_i|\mathbf{X}_i)$, then inclusion of the propensity scores variable merely causes the model to be over-fitted. If the nonresponse model is correctly specified, the propensity-related covariates remove the bias for estimating $E(Y_i|\mathbf{X}_i)$. It is true as long as the mean of Y_i varies smoothly with π_i and the relationship is arbitrarily well approximated by the linear combination of basis functions added to the complete-data model. This property is known as double robustness [11].

It should be noted that the logistic model may not accurately represent the true propensities for outliers (cases having $\hat{\pi}_i \approx 0$). If so, the inverse of the estimated propensities become larger, and the imputed values are inflated. Thus, the inverse propensity covariate (IPC) method might be sensitive to propensities near zero.

3.5. *Propensity dummy variable*

The dual modelling imputation model can be created in many ways. We can define a general framework of the dual modelling imputation model in the spirit of [19] that allows the mean response to vary with propensities in a flexible way. A simple form of an approximation of the relationship between the mean of y_i and $\hat{\pi}_i$ is a piecewise constant function at the sample quantiles of $\hat{\pi}_i$. This method, which has the double robustness property, is defined as follows: (a) classifies individuals into k classes in which the estimated propensity scores are nearly homogeneous; (b) creates $k - 1$ dummy variables to distinguish among the classes; (c) includes these dummy variables as predictors in the imputation model. In our simulation study, we consider $k = 5$. Kang and Schafer [10] proposed the same strategy for the mean estimation in presence of missing data.

4. Evaluation of the imputation methods

This section describes the evaluation of the performance of the five imputation methods described in the previous section via simulation. We used two data sets for generating incomplete data.

The first data set is the Irish wind speed data [20] which is used for continuous and dichotomous incomplete variables. It contains measurements of the average daily wind speed at 12 meteorological stations in Ireland for 18 years (1961–1978). High correlations between the stations, from 0.59 to 0.84, enable us to use the MAR mechanism that generates large differences between the complete and incomplete records. Here we are not interested in temporal variation between the measurements. The second data set is a study which was undertaken to assess factors associated with women's knowledge, attitude and behavioural towards mammography [21, p. 265]. The mammography experience (ME) has three categories (0 = never, 1 = during past year, 2 = over a year ago) which is suitable for a polytomous incomplete variable.

We applied the same setup as [22]. The missing data were generated under MCAR (missing completely at random) and different types of MAR mechanism: MARRIGHT, MARTAIL and MARMID. MARRIGHT removes large values in the distribution of the data. As an example, if Y is a continuous variable, a relatively large fraction of missing values occur in the large values of Y.

MARTAIL creates more missing values in both tails of this distribution. In other words, missing values are more likely with large or small values of Y. MARMID produces more missing values in the middle of the distribution which means that most missing values are located in the centre of the distribution of Y. Simulations were done using 1000 replications. All computations were done in R, v 2.11.1, using MICE (MICE, v 2.5; [23]). The results are presented in the following sections.

4.1. *Univariate continuous*

Five locations $(z_1, \dots, z_5)$ were selected from the Irish wind speed data. The original data z_1 were replaced by the newly generated data to avoid any issues of inaccuracy of model fit. A random sample of $n = 400$ was taken; 50% of the observations of z_1 were missing. In this case, the complete-data model was a linear regression $y_1 = z_1$ given covariates $\mathbf{x} = (z_2, \dots, z_5)$. A logistic regression was also used to calculate propensities for each individual in y_1. Covariates in the nonresponse model were defined by $\mathbf{u} = (z_2, \dots, z_5)$. The number of imputed data sets was 5. We chose the following complete-data statistics to investigate the validity of the imputation methods: the mean, the quartiles (the 25% quantile, median, and 75% quantile) and the Pearson correlation coefficient between the incomplete variable and the covariates. In addition, we calculated the coverage of a 95% confidence interval for each statistic.

Table 2 reports the simulation results. Except for propensity stratification (PS), all methods perform well. They produce unbiased estimates of the mean and the quartiles. The empirical coverage rates are very close to the nominal coverage (95%). In some cases, the coverage rate even exceeds the nominal level. The estimates of the Pearson correlations are unbiased with appropriate coverage rates. The PS method produces slightly biased estimates of the mean under MARMID. We observe biased estimates of the first quartile for all mechanisms even under MCAR. The estimate of the second quartile is also biased under MARRIGHT. Pearson's correlations are underestimated in all mechanisms. In addition, the coverage levels are less than or equal to 90% under MARRIGHT.

4.2. *Univariate dichotomous*

We selected five locations $(z_1, \dots, z_5)$ from the Irish wind speed data for the dichotomous case. Following [22], z_1 was dichotomized equally and then replaced with the newly generated data. Missing values were subsequently created for a sample of size 400. In this case, the relation between $y_1 = z_1$ and $\mathbf{x} = (z_2, \dots, z_5)$ was linked by a logistic function. Propensity scores were also estimated by a logistic regression with covariate $\mathbf{u} = (z_2, \dots, z_5)$. The same setup as the continuous case was used here. The quality of the five imputation methods was assessed by the proportions of each category and the conditional mean $\bar{x}$ within each category.

Table 3 presents the simulation results for the dichotomous incomplete variable. The probability of belonging to the first category $(y_1 = 0)$ is estimated correctly by the different methods, and the coverage rates achieve the nominal level. As in Table 2, we get biased estimates for PS. Under MARRIGHT, the IPC and propensity dummy variable (PDV) introduce some upward bias in the estimation of the conditional mean when $y_1 = 0$, but coverage rates remain in the acceptable region.

4.3. *Univariate polytomous*

We used the mammography data set to study the polytomous model. ME was the target variable, and the other five variables $(z_1, \dots, z_5)$ were the predictors of ME. The number of records in this

Table 2. Properties of different imputation methods in a continuous variable y_1 under MCAR, MARRIGHT, MARTAIL and MARMID.

				Propensity Scores		Dual modelling	
	Statistic	Pop	AP	PC	PS	IPC	PDV
MCAR	$E(y_1)$	11.66	11.66 (96)	11.66 (95)	11.68 (96)	11.66 (96)	11.66 (96)
	$P25(y_1)$	8.16	8.14 (98)	8.15 (98)	8.06 (97)	8.14 (98)	8.13 (98)
	$P50(y_1)$	11.42	11.40 (98)	11.40 (97)	11.39 (96)	11.40 (98)	11.40 (97)
	$P75(y_1)$	14.90	14.89 (98)	14.90 (98)	14.94 (97)	14.89 (98)	14.90 (98)
	$r(y_1 \cdot x_1)$	0.72	0.72 (94)	0.71 (95)	0.68 (92)	0.72 (94)	0.72 (94)
	$r(y_1 \cdot x_2)$	0.59	0.58 (95)	0.58 (95)	0.55 (95)	0.58 (95)	0.58 (95)
	$r(y_1 \cdot x_3)$	0.66	0.66 (95)	0.66 (94)	0.62 (94)	0.66 (94)	0.65 (95)
	$r(y_1 \cdot x_4)$	0.61	0.60 (94)	0.60 (94)	0.57 (93)	0.60 (94)	0.60 (94)
MARRIGHT	$E(y_1)$	11.66	11.68 (96)	11.67 (96)	11.70 (96)	11.68 (96)	11.68 (96)
	$P25(y_1)$	8.16	8.14 (97)	8.13 (97)	**7.80** (92)	8.13 (97)	8.11 (97)
	$P50(y_1)$	11.42	11.40 (97)	11.40 (97)	**11.13** (91)	11.39 (97)	11.38 (97)
	$P75(y_1)$	14.90	14.93 (98)	14.91 (97)	14.96 (97)	14.89 (97)	14.93 (98)
	$r(y_1 \cdot x_1)$	0.72	0.72 (94)	0.72 (93)	0.64 (**83**)	0.70 (94)	0.71 (93)
	$r(y_1 \cdot x_2)$	0.59	0.58 (93)	0.58 (94)	0.52 (90)	0.57 (94)	0.58 (93)
	$r(y_1 \cdot x_3)$	0.66	0.66 (94)	0.65 (95)	0.58 (**88**)	0.64 (94)	0.65 (94)
	$r(y_1 \cdot x_4)$	0.61	0.60 (95)	0.60 (95)	0.53 (90)	0.59 (94)	0.60 (93)
MARTAIL	$E(y_1)$	11.66	11.67 (96)	11.67 (96)	11.66 (96)	11.67 (96)	11.67 (95)
	$P25(y_1)$	8.16	8.16 (97)	8.15 (96)	8.06 (97)	8.15 (97)	8.14 (97)
	$P50(y_1)$	11.42	11.42 (97)	11.42 (97)	11.36 (96)	11.41 (97)	11.41 (97)
	$P75(y_1)$	14.90	14.89 (98)	14.90 (98)	14.89 (97)	14.90 (98)	14.89 (98)
	$r(y_1 \cdot x_1)$	0.72	0.72 (96)	0.72 (95)	0.67 (92)	0.72 (94)	0.71 (94)
	$r(y_1 \cdot x_2)$	0.59	0.58 (95)	0.58 (95)	0.54 (94)	0.58 (95)	0.58 (95)
	$r(y_1 \cdot x_3)$	0.66	0.66 (95)	0.66 (96)	0.61 (93)	0.66 (95)	0.65 (94)
	$r(y_1 \cdot x_4)$	0.61	0.60 (95)	0.60 (96)	0.56 (92)	0.60 (95)	0.60 (94)
MARMID	$E(y_1)$	11.66	11.67 (95)	11.67 (95)	**11.81** (96)	11.67 (95)	11.67 (95)
	$P25(y_1)$	8.16	8.14 (98)	8.14 (98)	8.06 (97)	8.13 (96)	8.11 (97)
	$P50(y_1)$	11.42	11.43 (98)	11.43 (98)	11.41 (97)	11.42 (98)	11.43 (97)
	$P75(y_1)$	14.90	14.93 (98)	14.93 (97)	14.99 (97)	14.92 (98)	14.95 (97)
	$r(y_1 \cdot x_1)$	0.72	0.72 (95)	0.72 (95)	0.69 (97)	0.72 (95)	0.72 (95)
	$r(y_1 \cdot x_2)$	0.59	0.58 (95)	0.58 (95)	0.56 (94)	0.58 (95)	0.58 (95)
	$r(y_1 \cdot x_3)$	0.66	0.66 (95)	0.66 (95)	0.63 (93)	0.66 (95)	0.66 (95)
	$r(y_1 \cdot x_4)$	0.61	0.60 (94)	0.60 (95)	0.57 (94)	0.60 (95)	0.60 (95)

Notes: $E(\cdot)$ represents the mean of y_1. $P25(y_1)$, $P50(y_1)$ and $P75(y_1)$ are used for the first, second and third quartiles of y_1. The Pearson correlation between y_1 and x is denoted by $r(y, x)$. The numbers represent the population value (Pop), and the average of the estimated quantity for the imputation methods is indicated by the name of the methods as follows: additional predictors (AP), propensity covariate (PC), propensity stratification (PS), propensity dummy variable (PDV), inverse propensity covariate (IPC); 95% confidence interval coverage is displayed in parentheses. Bias > 0.1 and coverage < 90 are given in bold.

data set was 412. The same setup was used as in Section 4.1. The complete-data model is a polytomous logistic regression with dependent variable ME ($= y_1$) and predictors $\mathbf{x} = (z_1, \ldots, z_5)$. The nonresponse model is also a logistic function of missing indicators depending on $\mathbf{u} = (z_1, \ldots, z_5)$.

Simulation results for the polytomous incomplete variable are shown in Table 4. The same statistics as the dichotomous incomplete variable are considered here. All methods perform well in estimating the marginal probability of observing the argument in y_1. The coverage levels do not show substantially different results from the nominal level except for the PS under MARRIGHT, where the coverage levels are relatively low. If we ignore the PS, the conditional expectation is estimated correctly under all missingness mechanisms. The PS does not perform well under MARRIGHT as shown by the highly biased estimates. In general, the coverage rates of all methods achieve the nominal level; sometimes the coverage rates are less than 90%, especially under MARRIGHT.

Table 3. Properties of different imputation methods in a dichotomous variable y_1 under MCAR, MARRIGHT, MARTAIL and MARMID.

				Propensity scores		Dual modelling	
	Statistic	Pop	AP	PC	PS	IPC	PDV
MCAR	$P(y_1=0)$	0.50	0.50 (95)	0.50 (95)	0.50 (96)	0.50 (95)	0.50 (96)
	$E(x_1\|y_1=0)$	8.59	8.61 (96)	8.61 (95)	**8.78** (95)	8.61 (95)	8.64 (95)
	$E(x_1\|y_1=1)$	16.17	16.16 (95)	16.16 (96)	**16.03** (95)	16.16 (95)	16.14 (95)
	$E(x_2\|y_1=0)$	9.24	9.24 (94)	9.24 (94)	9.36 (95)	9.25 (95)	9.27 (95)
	$E(x_2\|y_1=1)$	14.09	14.09 (96)	14.08 (96)	14.01 (94)	14.08 (95)	14.07 (96)
	$E(x_3\|y_1=0)$	7.09	7.12 (95)	7.12 (96)	**7.27** (94)	7.13 (96)	7.16 (96)
	$E(x_3\|y_1=1)$	13.84	13.84 (96)	13.84 (96)	**13.72** (94)	13.84 (96)	13.82 (96)
	$E(x_4\|y_1=0)$	7.12	7.13 (96)	7.13 (96)	**7.24** (96)	7.13 (96)	7.16 (96)
	$E(x_4\|y_1=1)$	12.50	12.50 (96)	12.50 (95)	11.46 (96)	12.50 (95)	12.48 (95)
MARRIGHT	$P(y_1=0)$	0.50	0.51 (96)	0.51 (96)	0.46 (**84**)	0.51 (96)	0.52 (96)
	$E(x_1\|y_1=0)$	8.59	8.67 (95)	8.68 (96)	**9.34** (**75**)	**8.82** (98)	**8.85** (97)
	$E(x_1\|y_1=1)$	16.17	16.17 (95)	16.14 (95)	**15.94** (95)	16.07 (95)	16.11 (95)
	$E(x_2\|y_1=0)$	9.24	9.29 (95)	9.30 (94)	**9.73** (**84**)	**9.38** (96)	**9.40** (96)
	$E(x_2\|y_1=1)$	14.09	14.08 (96)	14.06 (96)	**13.94** (95)	14.02 (96)	14.05 (96)
	$E(x_3\|y_1=0)$	7.09	7.16 (96)	7.17 (96)	**7.77** (**76**)	**7.29** (97)	**7.32** (97)
	$E(x_3\|y_1=1)$	13.84	13.84 (93)	13.82 (93)	**13.64** (93)	13.76 (94)	13.80 (95)
	$E(x_4\|y_1=0)$	7.12	7.18 (94)	7.18 (95)	**7.67** (**82**)	**7.28** (96)	**7.30** (95)
	$E(x_4\|y_1=1)$	12.50	12.49 (95)	12.47 (96)	**12.32** (94)	12.42 (96)	12.45 (96)
MARTAIL	$P(y_1=0)$	0.50	0.50 (95)	0.50 (96)	0.50 (95)	0.50 (95)	0.50 (95)
	$E(x_1\|y_1=0)$	8.59	8.63 (96)	8.63 (96)	**8.89** (92)	8.64 (95)	8.71 (96)
	$E(x_1\|y_1=1)$	16.17	16.14 (96)	16.14 (96)	**15.97** (93)	16.14 (96)	16.11 (96)
	$E(x_2\|y_1=0)$	9.24	9.26 (95)	9.26 (94)	**9.43** (95)	9.27 (94)	9.31 (96)
	$E(x_2\|y_1=1)$	14.09	14.09 (95)	14.09 (96)	**13.98** (95)	14.09 (95)	14.07 (95)
	$E(x_3\|y_1=0)$	7.09	7.14 (95)	7.14 (96)	**7.37** (92)	7.14 (96)	7.20 (96)
	$E(x_3\|y_1=1)$	13.84	13.81 (96)	13.81 (96)	**13.67** (92)	13.82 (95)	13.79 (96)
	$E(x_4\|y_1=0)$	7.12	7.15 (96)	7.14 (95)	**7.33** (93)	7.15 (96)	7.20 (96)
	$E(x_4\|y_1=1)$	12.49	12.48 (95)	12.48 (95)	**12.36** (95)	12.48 (95)	12.45 (95)
MARMID	$P(y_1=0)$	0.50	0.50 (96)	0.50 (95)	0.50 (96)	0.50 (97)	0.50 (96)
	$E(x_1\|y_1=0)$	8.59	8.60 (95)	8.61 (94)	**8.67** (95)	8.62 (95)	8.65 (95)
	$E(x_1\|y_1=1)$	16.17	16.19 (95)	16.19 (93)	16.07 (95)	16.19 (94)	16.17 (94)
	$E(x_2\|y_1=0)$	9.24	9.26 (95)	9.26 (96)	9.24 (95)	9.27 (95)	9.30 (95)
	$E(x_2\|y_1=1)$	14.09	14.11 (94)	14.11 (95)	14.03 (96)	14.10 (94)	14.09 (95)
	$E(x_3\|y_1=0)$	7.09	7.12 (94)	7.12 (95)	7.09 (95)	7.13 (95)	7.16 (94)
	$E(x_3\|y_1=1)$	13.84	13.87 (95)	13.87 (96)	13.74 (96)	13.87 (95)	13.85 (96)
	$E(x_4\|y_1=0)$	7.12	7.13 (94)	7.13 (94)	7.12 (95)	7.14 (95)	7.17 (95)
	$E(x_4\|y_1=1)$	12.49	12.52 (94)	12.52 (94)	12.42 (96)	12.51 (94)	12.50 (94)

Notes: $P(\cdot)$ is the marginal probability of observing the argument in y_1. $E(\cdot|\cdot)$ represents the conditional expectation of predictors. The numbers represent the population value (Pop), and the average of the estimated quantity for the imputation methods is indicated by the name of the methods as follows: additional predictors (AP), propensity covariate (PC), propensity stratification (PS), propensity dummy variable (PDV), inverse propensity covariate (IPC) along with their empirical coverage (in parentheses). Bias >0.1 and coverage <90 are given in bold.

4.4. *Multivariate continuous*

This section studies the performance of the five imputation methods for multivariate missing data. Six locations $(z_1, \ldots, z_6)$ were selected from the Irish wind speed data. We draw two data sets of size 400: one from the multivariate normal distribution to present an ideal case where there is no model misspecification and one from the raw data to present a practical situation that evaluates the robustness of the imputation methods.

Conditional on the observed data, four variables $\mathbf{y} = (z_1, \ldots, z_4)$ were subject to missingness (see [22] for details). The percentage of cases with missing data was 62.5. We used the Gibbs sampler technique [24] for the imputation process. The sampler consists of a set of linear regressions of

Table 4. Properties of different imputation methods in a polytomous variable y_1 under MCAR, MARRIGHT, MARTAIL and MARMID.

				Propensity scores		Dual modelling	
	Statistic	Pop	AP	PC	PS	IPC	PDV
MCAR	$P(y_1 = 0)$	0.25	0.25 (93)	0.25 (94)	0.26 (92)	0.25 (93)	0.26 (93)
	$P(y_1 = 1)$	0.18	0.18 (93)	0.18 (92)	0.19 (92)	0.18 (92)	0.18 (93)
	$P(y_1 = 2)$	0.57	0.57 (94)	0.57 (92)	0.55 (91)	0.57 (92)	0.56 (93)
	$E(x_1 \vert y_1 = 0)$	6.69	6.69 (94)	6.69 (94)	6.73 (94)	6.70 (93)	6.70 (95)
	$E(x_1 \vert y_1 = 1)$	7.19	7.21 (92)	7.21 (92)	7.22 (93)	7.21 (93)	9.21 (93)
	$E(x_1 \vert y_1 = 2)$	8.06	8.06 (99)	8.06 (99)	8.06 (99)	8.06 (99)	8.06 (98)
MARRIGHT	$P(y_1 = 0)$	0.25	0.26 (92)	0.26 (92)	0.27 (**80**)	0.26 (92)	0.26 (92)
	$P(y_1 = 1)$	0.18	0.18 (91)	0.18 (90)	0.20 (**78**)	0.18 (90)	0.18 (90)
	$P(y_1 = 2)$	0.57	0.56 (93)	0.56 (91)	0.53 (96)	0.56 (92)	0.56 (90)
	$E(x_1 \vert y_1 = 0)$	6.69	6.70 (**88**)	6.72 (**88**)	**6.91** (**66**)	6.75 (**87**)	6.74 (**87**)
	$E(x_1 \vert y_1 = 1)$	7.19	7.21 (**88**)	7.23 (**87**)	**7.45** (**69**)	7.26 (**86**)	9.24 (**83**)
	$E(x_1 \vert y_1 = 2)$	8.06	8.04 (99)	8.03 (99)	**7.90** (**86**)	8.01 (96)	8.02 (96)
MARTAIL	$P(y_1 = 0)$	0.25	0.26 (92)	0.26 (92)	0.26 (92)	0.26 (93)	0.26 (92)
	$P(y_1 = 1)$	0.18	0.18 (91)	0.18 (91)	0.19 (90)	0.18 (91)	0.18 (91)
	$P(y_1 = 2)$	0.57	0.56 (92)	0.56 (93)	0.55 (90)	0.56 (93)	0.56 (92)
	$E(x_1 \vert y_1 = 0)$	6.69	6.71 (92)	6.71 (92)	6.78 (90)	6.71 (92)	6.72 (91)
	$E(x_1 \vert y_1 = 1)$	7.19	7.21 (**89**)	7.20 (90)	7.27 (**88**)	7.20 (**88**)	9.20 (**88**)
	$E(x_1 \vert y_1 = 2)$	8.06	8.06 (96)	8.07 (96)	8.03 (96)	8.07 (96)	8.07 (96)
MARMID	$P(y_1 = 0)$	0.25	0.26 (93)	0.26 (94)	0.26 (91)	0.26 (94)	0.26 (92)
	$P(y_1 = 1)$	0.18	0.18 (92)	0.18 (92)	0.19 (90)	0.18 (92)	0.18 (91)
	$P(y_1 = 2)$	0.57	0.56 (94)	0.56 (94)	0.55 (90)	0.56 (93)	0.56 (91)
	$E(x_1 \vert y_1 = 0)$	6.69	6.69 (96)	6.69 (97)	6.75 (93)	6.70 (95)	6.71 (95)
	$E(x_1 \vert y_1 = 1)$	7.19	7.18 (95)	7.18 (95)	7.25 (93)	7.19 (94)	7.20 (95)
	$E(x_1 \vert y_1 = 2)$	8.06	8.05 (99)	8.06 (99)	8.03 (99)	8.06 (99)	8.06 (99)

Notes: The marginal probability of observing the argument and the conditional expectation of predictor variables are shown by $P(\cdot)$ and $E(\cdot|\cdot)$, respectively. The population value is given by (Pop), and the average of the estimated quantity for the imputation methods is also indicated by the name of the methods as follows: additional predictors (AP), propensity covariate (PC), propensity stratification (PS), propensity dummy variable (PDV), inverse propensity covariate (IPC). Empirical coverage levels are given in parentheses. Bias >0.1 and coverage <90 are given in bold.

each element **y** on all other variables. Propensity scores for each element of **y** were also calculated within the sampler. For instance, the propensity scores of z_1 were calculated using the variables $z_2, \ldots, z_6$ that were completed in the previous iteration of the Gibbs sampler. The number of Gibbs sampling iterations was 5, and the number of multiple imputed data sets was 10.

Table 5 shows the simulation results for the multivariate missing data under MARRIGHT. Estimates of the mean are correct with high coverage rates in both data sets. Quartiles are estimated properly, and the coverage rates are almost equal to the nominal level. The PS produces biased results for the quartiles, but the coverage rates remain at an acceptable level except for z_4. Estimates of the third quartile for z_1 and z_2 and the first quartile for z_4 are also affected by the missing data in all methods. Except for PS, the Pearson correlation is estimated accordingly everywhere with high coverage rates. We observe low coverage rates for the estimated Pearson correlation between z_1 and z_4 in the raw data.

4.5. *Conclusion*

Properties of the imputation methods were evaluated in a wide range of the complete-data statistics such as the mean, quartiles and Pearson correlation. We observed that all methods perform very well, except for PS. The PS method produces biased results especially when there is a dichotomous or polytomous incomplete variable. The IPC and PDV methods sometimes are less accurate in

Table 5. Properties of different imputation methods in multivariate continuous variables ($z_1, \ldots, z_4$) under MARRIGHT from two data sets (simulated and raw) based on Irish wind speed data.

	Simulated data						Raw data					
			Propensity scores		Dual modelling				Propensity scores		Dual modelling	
Statistic	Pop	AP	PC	PS	IPC	PDV	Pop	AP	PC	PS	IPC	PDV
$E(z_1)$	12.36	12.38 (96)	12.38 (95)	12.32 (98)	12.35 (94)	12.38 (96)	12.36	12.37 (96)	12.37 (94)	12.31 (98)	12.37 (95)	12.36 (96)
P25(z_1)	8.61	8.56 (97)	8.55 (97)	**8.47** (97)	8.53 (97)	8.55 (97)	8.12	8.17 (97)	8.16 (97)	8.06 (97)	8.17 (96)	8.16 (97)
P50(z_1)	12.40	12.36 (96)	12.35 (97)	**12.26** (94)	12.32 (96)	12.35 (96)	11.71	11.77 (98)	11.77 (97)	11.64 (96)	11.76 (96)	11.75 (97)
P75(z_1)	16.25	16.15 (97)	16.15 (97)	**16.05** (94)	**16.12** (95)	16.15 (97)	15.92	15.89 (98)	15.90 (98)	**15.75** (96)	15.88 (97)	15.87 (98)
$E(z_2)$	11.66	11.66 (95)	11.67 (93)	11.63 (98)	11.64 (94)	11.66 (96)	11.66	11.66 (95)	11.66 (95)	11.62 (98)	11.66 (96)	11.66 (96)
P25(z_2)	8.26	8.27 (98)	8.26 (97)	**8.14** (97)	8.24 (96)	8.26 (98)	8.00	7.91 (97)	7.90 (97)	**7.48** (96)	7.89 (96)	7.89 (97)
P50(z_2)	11.68	11.65 (96)	11.65 (97)	**11.57** (96)	11.61 (95)	11.64 (96)	10.92	11.02 (97)	11.00 (97)	10.93 (98)	10.99 (98)	10.99 (98)
P75(z_2)	15.16	**15.03** (96)	**15.04** (96)	**14.99** (95)	**15.00** (95)	**15.03** (97)	14.67	**14.80** (97)	**14.80** (97)	14.71 (97)	**14.78** (97)	14.77 (98)
$E(z_3)$	10.46	10.47 (95)	10.45 (96)	10.45 (98)	10.46 (95)	10.47 (95)	10.46	10.45 (96)	10.46 (94)	10.44 (98)	10.46 (95)	10.45 (97)
P25(z_3)	7.10	7.11 (97)	7.10 (97)	7.05 (97)	7.11 (96)	7.10 (96)	6.75	6.77 (97)	6.78 (96)	6.70 (96)	6.78 (97)	6.76 (97)
P50(z_3)	10.51	10.46 (96)	10.43 (95)	**10.38** (94)	10.43 (97)	11.45 (95)	9.96	9.97 (98)	9.96 (97)	**9.86** (97)	9.97 (98)	9.95 (98)
P75(z_3)	13.83	13.79 (97)	13.76 (97)	13.71 (96)	13.75 (96)	13.79 (97)	13.54	13.55 (98)	13.54 (97)	13.45 (97)	13.54 (98)	13.54 (98)
$E(z_4)$	9.79	9.80 (95)	9.81 (94)	9.76 (90)	9.79 (95)	9.80 (95)	9.79	9.79 (94)	9.80 (95)	9.76 (98)	9.80 (95)	9.79 (94)
P25(z_4)	6.54	**6.42** (96)	**6.41** (94)	**6.36** (**87**)	**6.40** (95)	**6.42** (96)	6.00	6.05 (96)	6.05 (96)	5.98 (98)	6.07 (97)	6.04 (97)
P50(z_4)	9.75	9.78 (97)	9.78 (96)	9.69 (**87**)	9.76 (96)	9.78 (97)	9.21	9.21 (97)	9.20 (96)	**9.07** (95)	9.22 (97)	9.19 (96)
P75(z_4)	13.14	13.15 (97)	13.15 (97)	13.13 (**85**)	13.13 (97)	13.15 (98)	12.96	12.92 (97)	12.93 (97)	**12.79** (96)	12.91 (97)	12.90 (97)
$r(z_1.z_2)$	0.72	0.72 (94)	0.72 (95)	0.61 (96)	0.72 (94)	0.72 (95)	0.72	0.71 (92)	0.72 (90)	0.59 (95)	0.71 (91)	0.71 (92)
$r(z_1.z_3)$	0.83	0.83 (95)	0.83 (95)	0.73 (96)	0.83 (95)	0.83 (95)	0.83	0.84 (90)	0.84 (88)	0.69 (97)	0.84 (91)	0.84 (93)
$r(z_1.z_4)$	0.73	0.73 (95)	0.73 (95)	0.64 (97)	0.73 (95)	0.73 (95)	0.73	0.77 (**76**)	0.77 (**73**)	0.62 (**80**)	0.77 (**81**)	0.77 (**80**)
$r(z_1.z_5)$	0.75	0.75 (95)	0.75 (95)	0.70 (97)	0.75 (94)	0.75 (95)	0.75	0.75 (93)	0.75 (94)	0.68 (98)	0.75 (96)	0.75 (94)
$r(z_1.z_6)$	0.62	0.62 (95)	0.62 (95)	0.58 (97)	0.62 (94)	0.62 (96)	0.62	0.63 (92)	0.63 (93)	0.57 (98)	0.63 (94)	0.63 (93)
$r(z_2.z_3)$	0.59	0.58 (94)	0.58 (95)	0.50 (97)	0.58 (95)	0.58 (95)	0.59	0.60 (94)	0.60 (94)	0.49 (98)	0.60 (95)	0.60 (94)
$r(z_2.z_4)$	0.66	0.66 (95)	0.66 (95)	0.57 (97)	0.66 (94)	0.65 (95)	0.66	0.68 (92)	0.68 (91)	0.55 (97)	0.68 (92)	0.67 (92)
$r(z_2.z_5)$	0.61	0.60 (96)	0.61 (96)	0.55 (97)	0.60 (94)	0.60 (96)	0.61	0.59 (92)	0.60 (92)	0.54 (97)	0.60 (94)	0.60 (93)
$r(z_2.z_6)$	0.47	0.47 (95)	0.48 (95)	0.43 (97)	0.48 (94)	0.47 (96)	0.47	0.46 (91)	0.47 (91)	0.43 (97)	0.47 (93)	0.46 (93)
$r(z_3.z_4)$	0.79	0.79 (95)	0.79 (96)	0.70 (97)	0.78 (96)	0.78 (96)	0.79	0.78 (93)	0.78 (93)	0.67 (97)	0.78 (93)	0.78 (93)
$r(z_3.z_5)$	0.82	0.82 (95)	0.82 (95)	0.77 (97)	0.82 (96)	0.82 (96)	0.82	0.82 (94)	0.82 (92)	0.75 (98)	0.82 (95)	0.82 (95)
$r(z_3.z_6)$	0.67	0.67 (94)	0.67 (94)	0.63 (98)	0.67 (95)	0.67 (95)	0.67	0.67 (93)	0.68 (94)	0.62 (98)	0.67 (94)	0.67 (93)
$r(z_4.z_5)$	0.84	0.84 (95)	0.84 (96)	0.80 (97)	0.84 (95)	0.84 (96)	0.84	0.85 (93)	0.85 (90)	0.78 (98)	0.85 (92)	0.85 (94)
$r(z_4.z_6)$	0.77	0.77 (94)	0.77 (94)	0.73 (98)	0.77 (96)	0.76 (95)	0.77	0.77 (92)	0.77 (92)	0.71 (98)	0.77 (93)	0.77 (92)

Notes: The mean and quartiles for each incomplete variable are shown as $E(\cdot)$, P25($\cdot$), P50($\cdot$) and P75($\cdot$). r($\cdot,\cdot$) represents the Pearson correlation between variables. Variables z_5 and z_6 are complete. The numbers represent the population value (Pop), and the average of the estimated quantity for the imputation methods is indicated by the name of the methods as follows: additional predictors (AP), propensity covariate (PC), propensity stratification (PS), propensity dummy variable (PDV), inverse propensity covariate (IPC); 95% confidence interval coverage is displayed in parentheses. Bias >0.1 and coverage <90 are given in bold.

the MARRIGHT mechanism. The results are also very similar for additional predictor (AP) and propensity covariate (PC) methods.

5. Discussion

It has been widely acknowledged that the imputation model should incorporate predictors that appear in the complete-data model and predictors that are related to the nonresponse model. It is not clear, however, how these two types of predictors should be combined into one model to generate the imputed data. This paper investigated the properties of five different combination methods. Under MAR the results of the five methods were similar, although the PS was less precise. This means that, except for PS, there is no *a priori* preference between the four other methods based on their statistical properties.

The additive predictors imputation method is simple to use because it consists of just adding extra variables to the imputation model. However, a data set may contain lots of variables, say hundreds or thousands, which may lead to models that are too large. Including all variables into the imputation model may therefore not always be the best solution.

When the number of variables grows, we may find that the ideal model is too large to implement. The imputation method based on the propensity scores has an attractive feature, which is the ability of replacing a potentially large set of variables in the nonresponse model with a single aggregate of these variables. However, imputation methods using propensity scores are not without limitations. For instance, logistic regression is often used to estimate propensity scores, but may be unrealistic for the data at hand.

Another approach is the dual modelling imputation method, which can be implemented in several ways. We investigated the use of the inverse of the propensity scores. Like the propensity scores method, the inverse propensity scores method is sensitive to misspecification of the propensity model especially when the propensity scores are close to zero. Therefore, care should be taken to select an estimator that is not sensitive to the misspecification of the nonresponse model.

Here, we assumed both the complete-data and the nonresponse models were correctly specified. But, in practice, it is hard to discover whether the models are correct. We considered different variants of the MAR mechanism. Although MAR is highly useful as an initial assumption, it may be dubious for certain applications. We do not yet know if our results generalize to missing not at random.

This paper emphasizes the combination of the complete-data model and the nonresponse model. Four methods out of five perform equally well and the PS method is not recommended. From the four methods, AP is the easiest to implement. In short, we underline that it *does matter* which variables are included in the imputation model, but it *does not matter* so much how to do this.

Acknowledgements

The authors are truly grateful to the reviewers for their useful comments that helped improving the paper.

References

[1] O. Harel and X.H. Zhou, *Multiple imputation: Review of theory, implementation and software*, Stat. Med. 26 (2007), pp. 3057–3077.
[2] M.G. Kenward and J. Carpenter, *Multiple imputation: Current perspectives*, Stat. Methods Med. Res. 16 (2007), pp. 199–218.
[3] P. Zhang, *Multiple imputation: Theory and method*, Int. Stat. Rev. 71 (2003), pp. 581–592.
[4] D.B. Rubin, *Multiple Imputation for Nonresponse in Surveys*, John Wiley, New York, 1987.

[5] L.M. Collins, J.L. Schafer, and C.M. Kam, *A comparison of inclusive and restrictive strategies in modern missing data procedures*, Psychol. Methods 6 (2001), pp. 330–351.
[6] S. van Buuren, H.C. Boshuizen, and D.L. Knook, *Multiple imputation of missing blood pressure covariates in survival analysis*, Stat. Med. 18 (1999), pp. 681–694.
[7] J.P Reiter, T.E. Raghunathan, and S.K. Kinney, *The importance of modeling the sampling design in multiple imputation for missing data*, Surv. Methodol. 32 (2006), pp. 143–149.
[8] M. Spratt, J. Carpenter, J.A.C. Sterne, J.B. Carlin, J. Heron, J. Henderson, and K. Tilling, *Strategies for multiple imputation in longitudinal studies*, Am. J. Epidemiol. 172 (2010), pp. 478–487.
[9] P.W. Lavori, R. Dawson, and D. Shera, *Multiple imputation strategy for clinical trials with truncation of patient data*, Stat. Med. 14 (1995), pp. 1913–1925.
[10] J.D.Y. Kang and J.L. Schafer, *Demystifying double robustness: A comparison of alternative strategies for estimating population means from incomplete data*, Statist. Sci. 26 (2007), pp. 523–539.
[11] H. Bang and J.M. Robins, *Doubly robust estimation in missing data and causal inference models*, Biometrics 61 (2005), pp. 962–972.
[12] D.B. Rubin, *Inference and missing data*, Biometrika 63 (1976), pp. 581–592.
[13] P.R. Rosenbaum and D.B. Rubin, *The central role of the propensity score in observational studies for causal effects*, Biometrika 70 (1983), pp. 41–55.
[14] D.O. Scharfstein, A. Rotnitzky, and J.M. Robins, *Adjusting for nonignorable drop-out using semiparametric nonresponse models*, J. Am. Statist. Assoc. 94 (1999), pp. 1096–1120 (with Rejoinder, pp. 1135–1146).
[15] J.M. Robins, *Robust estimation in sequentially ignorable missing data and causal inference models*, Proceedings of the American Statistical Association Section on Bayesian Statistical Science, Alexandria, VA, 2000, pp. 6–10.
[16] M.J. van der Laan and J.M. Robins, *Unified Methods for Censored Longitudinal Data and Causality*, Springer Verlag, New York, 2003.
[17] R. Neugebauer and M.J. van der Laan, *Why prefer double robust estimates?* J. Stat. Plann. Inference 129 (2005), pp. 405–426.
[18] J. Carpenter, M. Kenward, and S. Vansteelandt, *A comparison of multiple imputation and doubly robust estimation for analyses with missing data*, J. R. Statist. Soc. A 169 (2006), pp. 571–584.
[19] R. Little and H. An, *Robust likekihood-based analysis of multivariate data with missing values*, Statist. Sinica 14 (2004), pp. 949–969.
[20] J. Hasslet and A.E. Raftery, *Space-time modeling with long-memory dependence: Assessing Ireland's wind power resource*, Appl. Stat. 38 (1989), pp. 1–50.
[21] D.W. Hosmer and S. Lemeshow, *Applied Logistic Regression*, John Wiley, New York, 2003.
[22] S. van Buuren, J.P.L. Brand, C.G.M. Groothuis-Oudshoorn, and D.B. Rubin, *Fully conditional specification in multivariate imputation*, J. Stat. Comput. Simul. 76 (2006), pp. 1048–1064.
[23] S. van Buuren and K.G. Oudshoorn, *MICE: Multivariate Imputation by Chained Equations in R*, J. Stat. Softw., forthcoming.
[24] A.E. Gelfand, *Gibbs sampling*, J. Am. Statist. Assoc. 95 (2000), pp. 1300–1304.