British Journal of Mathematical and Statistical Psychology (2013)

The British Psychological Society

www.wileyonlinelibrary.com

Dual imputation model for incomplete longitudinal data

Shahab Jolani[1]*, Laurence E. Frank[1] and Stef van Buuren[1,2]

[1]Department of Methodology and Statistics, Utrecht University, Utrecht, The Netherlands

[2]Department of Statistics, TNO, Leiden, The Netherlands

Missing values are a practical issue in the analysis of longitudinal data. Multiple imputation (MI) is a well-known likelihood-based method that has optimal properties in terms of efficiency and consistency if the imputation model is correctly specified. Doubly robust (DR) weighing-based methods protect against misspecification bias if one of the models, but not necessarily both, for the data or the mechanism leading to missing data is correct. We propose a new imputation method that captures the simplicity of MI and protection from the DR method. This method integrates MI and DR to protect against misspecification of the imputation model under a missing at random assumption. Our method avoids analytical complications of missing data particularly in multivariate settings, and is easy to implement in standard statistical packages. Moreover, the proposed method works very well with an intermittent pattern of missingness when other DR methods can not be used. Simulation experiments show that the proposed approach achieves improved performance when one of the models is correct. The method is applied to data from the fireworks disaster study, a randomized clinical trial comparing therapies in disaster-exposed children. We conclude that the new method increases the robustness of imputations.

1. Introduction

An analysis of longitudinal data might suffer from missing values. In a clinical trial study, for instance, some intended measurements cannot be taken owing to unwillingness of the individuals. Failure to take missing data into account can lead to biased inferences.

One common assumption in the analysis of missing data is that they are missing at random (MAR) in the sense that the process governing missingness does not depend on the unobserved measurements, after conditioning on the observed measurements (Rubin, 1976). Verification of the MAR assumption using the observed data is not possible (Molenberghs, Beunckens, Sotto, & Kenward, 2008) so the plausibility of MAR must be thoughtfully considered using the expert knowledge or other external sources. In most practical cases, however, the MAR assumption is often assumed and found to be a more plausible assumption to start with (Schafer, 1997). Our best hope is that the MAR assumption nearly holds, and the impact of unobserved measurements on our conclusions is minimal. If it is suspected that the MAR assumption is violated, sensitivity analysis can be performed to investigate the limitation of modelling assumptions (see, for example, Kenward, 1998; Robins, Rotnitzky, & Scharfstein, 1999). Throughout this paper

*Correspondence should be addressed to Shahab Jolani, Department of Methodology and Statistics, Utrecht University, Padualaan 14, 3584 CH Utrecht, The Netherlands (e-mail: s.jolani@uu.nl).

DOI:10.1111/bmsp.12021

we assume either that the MAR assumption is plausible, or that we wish to conduct an MAR analysis as a point of departure for sensitivity analyses.

Multiple imputation (MI) is an accepted method of dealing with missing data. The use of MI is becoming increasingly routine because of its simplicity and software development (Schafer & Graham, 2002). The doubly robust (DR) method is an alternative approach that uses the augmented inverse probability weights (AIPWs) to refine the estimates of the parameters (e.g., Bang & Robins, 2005; Rotnitzky, 2009). The method requires specification of two models: one that models the distribution of the complete data, including both the outcomes and covariates, and another that models the missingness mechanism. If one of the two models is correctly specified, then the estimates of the parameters of interest are asymptotically unbiased.

Several authors have compared MI and DR (e.g., Carpenter, Kenward, & Vansteelandt, 2006; Jolani, van Buuren, & Frank, 2013; Qi, Wang, & He, 2010). The major advantage of MI is that standard complete data methods can be used to analyse the data. In other words, if the missing values are replaced with the imputed values, any statistical analyses can be applied to the completed data. However, MI enables valid inferences only if the imputation model is correctly specified. In contrast, the DR method provides estimates that are consistent if either the complete data model (the model of scientific interest) or the model for the missingness mechanism is correct. The method provides the analyst two chances of obtaining valid inferences, rather than one. The technical nature of the DR method, on the other hand, hampers practical application. Moreover, the DR method can be unstable in practice when both models are misspecified or when the probability weights are close to zero (Kang & Schafer, 2007). A number of authors have reported improved DR estimators concerning these issues (Cao, Tsiatis, & Davidian, 2009; Tsiatis, Davidian, & Cao, 2011).

Scharfstein, Rotnitzky, & Robins, (1999, pp. 1140–1141) showed the regression representation of the AIPW estimator in missing data models under MAR. This estimator is DR and can be constructed by including the inverse of the probability weights into the complete data model as an additional predictor (Vansteelandt, Carpenter, & Kenward, 2010). This methodology was further extended to longitudinal monotone missing data (Bang & Robins, 2005). A limitation of this extension is the lack of generalization to intermittent missing data in which several variables have missing observations in an arbitrary pattern. No general rule exists to extend their works in intermittent missing data. Recently, Daniel and Kenward (2012) proposed a DR-type imputation method based on the inverse of the probability weights that works for a subset of MAR, but it is not obvious how to extend their method to the imputation of multivariate missing data in general. In addition, propensities near zero produce extremely large inverses of the probability weights that cause explosion of the imputation values (Kang & Schafer, 2007; Vansteelandt *et al.*, 2010; Zhang & Little, 2011).

As the specification of the imputation model plays an important role in the validity of MI, it would be desirable if the imputation model could be protected against misspecification. To achieve this goal, we need to construct an imputation model that has the double robustness property. A general strategy to achieve this is to incorporate functions of the propensity score, the probability of being observed given some covariates, into the imputation model as a predictor.

The present paper introduces a new class of MI which is doubly protected, and which can be applied to any patterns of missingness including intermittent missing data. We integrate DR and MI, and thus construct a *dual imputation model* (DIM). The new method is designed to increase the robustness of the imputation model while attempting to avoid bias from the final inference.

The key feature of the DIM strategy is to iteratively estimate propensities for each incomplete variable conditional on the other variables, and to impute missing values on that variable by including a function of propensities (e.g., the inverse of the propensities) into the imputation model. The new imputation method is expected to be readily robust, and it is designed to solve problems with any pattern of missing data.

In the next section, we describe data from the Enschede fireworks disaster study and motivate the analysis of these data by the new imputation method. Section 3 describes a model that is typically used in longitudinal studies. Section 4 provides a general DIM framework when the missing data have an arbitrary pattern of missingness. The final sections evaluate the performance of our method in terms of simulation studies, and analysis of our data.

2. Motivating example

To help motivate the material in this paper, we briefly describe the fireworks disaster data (FDD; van Buuren, 2012). The fireworks disaster study was a randomized clinical trial to compare cognitive behavioural therapy (CBT) and eye movement desensitization and reprocessing (EMDR) in disaster-exposed children aged 4–18 years. This study was conducted after an explosion at a fireworks factory in Enschede, the Netherlands, in 2000. Fifty-two children and their parents, who showed signs of post-traumatic stress disorder (PTSD), were randomized to one of two treatment arms, CBT and EMDR (De Roos *et al.*, 2011). The primary outcome measure was the University of California at Los Angeles PTSD reaction index (PTSD-RI) obtained at pre-treatment, post-treatment (4–8 weeks), and follow-up (3 months) along with baseline covariates such as age and sex. The central goal was to compare the effectiveness of CBT and EMDR by comparing the means of the PTSD-RI in both treatments. We here compare the two treatment conditions by estimating the difference between the means of the PTSD-RI at the end of study. This provides marginal effects of the treatments on reducing PTSD-RI.

This study suffered from intermittent non-response. The reasons for skipping one or more occasions were that the parents were overburdened, refused to talk, or language problems. We assume the MAR mechanism is a plausible assumption in our data set, or at least it is considered as a suitable point of departure for conducting sensitivity analysis. Of the 52 children, 31, 42, and 46% had a missing measurement of the PTSD-RI per visit. The intermittent pattern of missingness clearly complicates statistical inferences. According to the intention-to-treat principle in clinical trials, the outcomes should be analysed using data from all randomized participants, which in this case is impossible because of the missing data.

A large number of variables exist in the FDD, and setting up an appropriate imputation model requires some care. Our plan is to apply the DIM strategy to reduce the bias, if any, in the imputation model. The propensity scores, if correctly estimated, can remove the bias for estimating the mean of the outcome variable (i.e., the mean of PTSD-RI). In sum, robustness of the imputed values might be increased by including the propensity related covariate in the imputation model.

3. Model

In a longitudinal study, it is often of interest to evaluate the mean of an outcome variable over time. Let Y_{it} be the measurement of the outcome variable for individuals $i = 1,\ldots,n$ at

times $t = 1,\ldots,T$. We assume that individuals are independent, and Y_{it} is a function of explanatory variables $\mathbf{X}_{it} = (\mathbf{X}_{i0}, \mathbf{X}_{it}^{*})'$. Note that a boldface letter indicates a vector throughout. In randomized studies, $\mathbf{X}_{i0}$ typically includes the treatment arm indicator and baseline characteristics such as age and sex, and $\mathbf{X}_{it}^{*}$ includes time-dependent covariates such as interactions of the baseline variables with time. We consider a longitudinal regression model where the conditional mean of Y_{it} given $\mathbf{X}_{it}$ can be specified as

$$E(Y_{it}|\mathbf{X}_{it}) = g(\mathbf{X}_{it};\boldsymbol{\beta}), \tag{1}$$

where $g(\mathbf{X}_{it};\boldsymbol{\beta})$ is a known function of $\boldsymbol{\beta}$, and $\boldsymbol{\beta}$ is a $p \times 1$ vector of unknown parameters in which we are interested. The regression model in equation (1) is flexible since it allows for the dependence of the mean of Y_{it} on baseline and time-dependent covariates. As an example, in the FDD, one can specify a linear regression $g(\mathbf{X}_{it};\boldsymbol{\beta}) = \beta_0 + \beta_1 t + \beta_2 A_i + \beta_3 A_i t$ with A_i defining the treatment arm indicator (CBT or EMDR), and $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, \beta_3)'$ containing the parameters of the model for the conditional expectation of Y_{it} given $\mathbf{X}_{it}$ which changes linearly with time, treatment arm indicator, and interaction between time and treatment arm indicator.

Further assume that, in addition to Y_{it} and $\mathbf{X}_{it}$, a vector of other time-dependent covariates $\mathbf{U}_{it}$ is also available at each occasion, but of no substantive interest in the sense that the scientific aim is the estimation of the conditional mean of Y_{it} given $\mathbf{X}_{it}$ rather than the conditional mean of Y_{it} given $\mathbf{X}_{it}$ and $\mathbf{U}_{it}$. In the FDD, for instance, the PTSD-RI scores of the parents are of secondary interest, and the effect of the randomized treatment on the PTSD-RI score of child i at time t, Y_{it}, would be reflected in the model in equation (1), which does not further adjust for the score of his or her parents ($\mathbf{U}_{it}$).

Suppose that $\mathbf{X}_{it}$ and $\mathbf{U}_{it}$ are always observed, but Y_{it} can be observed or missing. Let R_{it} denote a binary response indicator for individual i at time t; that is, $R_{it} = 1$ if Y_{it} is observed and $R_{it} = 0$ otherwise. We further assume that the probability of response follows a specified parametric model. In what follows it is convenient to define $\mathbf{Y}_{i(-t)} = (Y_{i1},\ldots,Y_{i(t-1)},Y_{i(t+1)},\ldots,Y_{iT})'$, and $\mathbf{W}_{it} = (\mathbf{Y}_{i(-t)},\mathbf{X}_{it},\mathbf{U}_{it})'$. Note that, for individual i at time t, $\mathbf{W}_{it}$ contains the outcome variable for all visits except for time t and full collections of variables $\mathbf{X}_{it}$ and $\mathbf{U}_{it}$.

Given $\mathbf{W}_{it}$, let r_{it} denote the realization of R_{it} for individual i at time t. The conditional probability $P(R_{it} = r_{it}|\mathbf{W}_{it})$ follows a parametric model up to a $q \times 1$ parameter vector $\boldsymbol{\alpha}_t$. Let $\pi_{it}(\boldsymbol{\alpha}_t) = P(R_{it} = 1|\mathbf{W}_{it})$; then

$$\text{logit}\{\pi_{it}(\boldsymbol{\alpha}_t)\} = h(\mathbf{W}_{it};\boldsymbol{\alpha}_t), t = 1,\ldots,T, \tag{2}$$

where $h(\cdot;\cdot)$ is a known function. A common candidate for the transformation of $\pi_{it}(\boldsymbol{\alpha}_t)$ in model in equation (2) is the logistic function. However, other transformations such as the probit function can be accommodated. It should be noted that the so-called propensity score model in equation (2) is a general model for the missingness mechanism. For example, in the FDD, one can consider $\text{logit}\{\pi_{it}(\boldsymbol{\alpha}_t)\} = \alpha_{0t} + \alpha_{1t}A_i + \alpha_{2t}U_{it} + \alpha_{3t}Y_{i(t-1)}$ where the mechanism of missingness at time t is a function of the treatment arm, the score of PTSD-RI for the parents at time t, and the score of PTSD-RI for the child at time $t-1$.

It is very important to recognize that the probability of being observed for Y_{it} in model in equation (2) does not depend on Y_{it}, while it can depend on the outcomes at the other time points (i.e., $Y_{i1},\ldots,Y_{i(t-1)},Y_{i(t+1)},\ldots,Y_{iT}$). This makes the model for the response probabilities (equation (2)) different from the missing not at random approaches where

the response probability for Y_{it} depends on the outcome variable Y_{it} as well as the other outcomes (see, for example, Diggle & Kenward, 1994).

4. Dual imputation model

4.1. Univariate missing data

In this section, we show how to construct a DIM for longitudinal data when the missing-data mechanism is ignorable (Little & Rubin, 2002, p. 119). Suppose, for simplicity, that only the outcome variable for the last occasion (Y_T) is incomplete. Let $Y_{T,\text{obs}}$ and $Y_{T,\text{mis}}$ denote the observed and missing parts of Y_T, respectively. Suppressing i from the notation, standard MI suggests (i) fitting a model $f(Y_T|\mathbf{X}_T, \mathbf{Y}_{-T}; \boldsymbol{\theta}_T)$ to the observed part of Y_T by, for instance, least squares, and drawing $\tilde{\boldsymbol{\theta}}_T$ from the posterior distribution $P(\boldsymbol{\theta}_T|\mathbf{X}_T, Y_{T,\text{obs}}, \mathbf{Y}_{-T})$, and (ii) imputing $Y_{T,\text{mis}}$ by the regression estimator $f(Y_T|\mathbf{X}_T, \mathbf{Y}_{-T}; \tilde{\boldsymbol{\theta}}_T)$ plus random noises.

The imputation model of the above case might suffer from misspecification especially when the uncertainty due to the model specification grows relative to sampling variation. Therefore, we attempt to address misspecification of the imputation model, if any, by making a modification in the direction of the propensity score π_T. This is motivated by the fact that the propensity score is a summary of $\mathbf{W}_T$ that makes Y_T and R_T conditionally independent (Rosenbaum, 2002).

Following Bang and Robins (2005), we incorporate the inverse of π_T into the imputation model as follows. We estimate π_T from the parametric model in equation (2) with a known link function (e.g., logit link), and add $\hat{\pi}_T^{-1}$ as an additional predictor into the imputation model. Note that inclusion of $\hat{\pi}_T^{-1}$ is a sufficient conditions to obtain a DR estimator in this setting (Scharfstein *et al.*, 1999). The new imputation model then fits the model $f(Y_T|\mathbf{X}_T, \mathbf{Y}_{-T}, \hat{\pi}_T^{-1}; \boldsymbol{\theta}_T^*) = f(Y_T|\mathbf{X}_T, \mathbf{Y}_{-T}; \boldsymbol{\theta}_T) + \gamma_T \hat{\pi}_T^{-1}$ to the observed part of Y_T where $\boldsymbol{\theta}_T^* = (\boldsymbol{\theta}_T, \gamma_T)$, and γ_T is a regression coefficient for the new predictor $\hat{\pi}_T^{-1}$. As an example, the imputation model can take the linear form $f(Y_T|\mathbf{X}_T, \mathbf{Y}_{-T}, \hat{\pi}_T^{-1}; \boldsymbol{\theta}_T^*) = \boldsymbol{\theta}_1' \mathbf{X}_T + \boldsymbol{\theta}_2' \mathbf{Y}_{-T} + \gamma_T \hat{\pi}_T^{-1}$, where $\boldsymbol{\theta}_T^* = (\boldsymbol{\theta}_1', \boldsymbol{\theta}_2', \gamma_T)$. A random draw $\tilde{\boldsymbol{\theta}}_T^*$ is made from its posterior distribution, and the missing part of Y_T is imputed under the regression estimator $f(Y_T|\mathbf{X}_T, \mathbf{Y}_{-T}, \hat{\pi}_T^{-1}; \tilde{\boldsymbol{\theta}}_T^*)$. It should be noted that the resulting imputations are doubly protected in the sense that either the model $f(Y_T|\mathbf{X}_T, \mathbf{Y}_{-T}; \boldsymbol{\theta}_T)$ or the model π_T is correct (Bang & Robins, 2005).

4.2. Multivariate missing data

Extending the DIM to incomplete longitudinal data with an arbitrary pattern of missingness is not straightforward. The obvious reason is the multivariate nature of missing data which creates complexity for making inferences. To describe the procedure, we should first consider a joint probability distribution for the complete responses. In a longitudinal study, when the repeated measurements are quantitative, it is common to assume a multivariate normal distribution for the correlated responses. Even if the assumption of normality might not ultimately be a realistic choice for other situations (e.g., binary cases), it may still be reasonable to assume a multivariate normal distribution for the purpose of imputation (Bernaards, Belin, & Schafer, 2007; Demirtas, Freels, & Yucel, 2008; Schafer, 1997).

More formally, write $\mathbf{Y} = \{Y_{it}\}$ and $\mathbf{R} = \{R_{it}\}$ as $n \times T$ matrices of outcomes and response indicators. We assume that the rows of $\mathbf{Y}$ are independent and identically

distributed from the joint distribution $P(\mathbf{Y};\boldsymbol{\theta})$ where we index the probability density function of $\mathbf{Y}$ by $\boldsymbol{\theta}$. Denote the observed part of $\mathbf{Y}$ by $\mathbf{Y}_{\text{obs}}$, and the missing part by $\mathbf{Y}_{\text{mis}}$, so that $\mathbf{Y} = (\mathbf{Y}_{\text{obs}}, \mathbf{Y}_{\text{mis}})$.

Under ignorability, the missingness mechanism formally plays no role in making inferences about $\boldsymbol{\theta}$. However, many researchers recognize that modelling the missingness mechanism is useful for model validation and criticism (Gelman, Carlin, Stern, & Rubin, 2004, ch. 6–7). Moreover, using such information for the purpose of imputation may lead to sharper inferences (Schafer, 2003). As the actual observed data are $\mathbf{V} = (\mathbf{Y}_{\text{obs}}, \mathbf{R}, \mathbf{X}, \mathbf{U})$, the posterior predictive distribution of missing data given the observed data can be defined as

$$P(\mathbf{Y}_{\text{mis}}|\mathbf{V}) = \int\int P(\mathbf{Y}_{\text{mis}}|\mathbf{V}, \boldsymbol{\theta}, \boldsymbol{\alpha}) p(\boldsymbol{\theta}, \boldsymbol{\alpha}|\mathbf{V}) \partial\boldsymbol{\theta}\partial\boldsymbol{\alpha}, \tag{3}$$

where $P(\boldsymbol{\theta},\boldsymbol{\alpha}|\mathbf{V}) \propto P(\mathbf{V}|\boldsymbol{\theta},\boldsymbol{\alpha}) p(\boldsymbol{\theta},\boldsymbol{\alpha})$, and $p(\boldsymbol{\theta},\boldsymbol{\alpha})$ is the prior density. Here, in fact, the imputation procedure proposed by Rubin (1987, p. 161) is extended by incorporating a nuisance parameter α into it. The purpose of including α is to increase robustness of imputations.

The ignorable missing data mechanism implies that the parameters $\boldsymbol{\theta}$ and $\boldsymbol{\alpha}$ are *a priori* independent, $p(\boldsymbol{\theta},\boldsymbol{\alpha}) = p(\boldsymbol{\theta})p(\boldsymbol{\alpha})$. The posterior distribution of the parameters $\boldsymbol{\theta}$ and $\boldsymbol{\alpha}$ given the observed data can then be written as $P(\boldsymbol{\theta}, \boldsymbol{\alpha}|\mathbf{V}) = P(\boldsymbol{\alpha}|\mathbf{X}, \mathbf{U}, \mathbf{Y}_{\text{obs}}, \mathbf{R}) P(\boldsymbol{\theta}|\mathbf{X}, \mathbf{Y}_{\text{obs}})$, where $P(\boldsymbol{\alpha}|\mathbf{X}, \mathbf{U}, \mathbf{Y}_{\text{obs}}, \mathbf{R}) \propto P(\mathbf{R}|\mathbf{X}, \mathbf{U}, \mathbf{Y}_{\text{obs}}; \boldsymbol{\alpha}) p(\boldsymbol{\alpha})$ and $P(\boldsymbol{\theta}|\mathbf{X}, \mathbf{Y}_{\text{obs}}) \propto P(\mathbf{Y}_{\text{obs}}|\mathbf{X}; \boldsymbol{\theta}) p(\boldsymbol{\theta})$. In this paper, we use standard non-informative priors for parameters $\boldsymbol{\theta}$ and $\boldsymbol{\alpha}$.

Direct simulation from equation (3) is difficult due to the multivariate nature of missing data and general pattern of missingness. One standard solution is to draw approximate samples using a Gibbs-type sampling scheme. That is, the missing values in each incomplete variable are imputed conditional on the other variables. In what follows, we suppress i from the notation. For each incomplete variable Y_t, $t = 1,\ldots,T$, the model in equation (2) can be used to draw a random value of α and estimate the propensity score π_t^{-1} based on the drawn value. The imputation model $f(Y_t|\mathbf{X}_t, \mathbf{Y}_{-t}, \pi_t^{-1})$ then generates imputations for the missing part of Y_t. Note that we include π_t^{-1} in a semi-parametric fashion in the imputation model as it is a necessary condition to obtain a DR estimator (Bang & Robins, 2005). Cycling through all the models, posterior draws of the parameters are made given current values of the other variables. More specifically, one can draw samples through the following steps at iteration k:

Step 1. Draw $\boldsymbol{\alpha}_1^{(k)}$ from $P(\boldsymbol{\alpha}_1|\mathbf{W}_1^{(k-1)}, R_1)$; draw $\boldsymbol{\theta}_1^{(k)}$ from $P(\boldsymbol{\theta}_1|\mathbf{X}_1, Y_{1,\text{obs}}, \mathbf{Y}_{-1}^{(k-1)})$; draw $Y_{1,\text{mis}}^{(k)}$ from $P(Y_1|\mathbf{X}_1, \mathbf{Y}_{-1}^{(k-1)}, \boldsymbol{\theta}_1^{(k)}, \boldsymbol{\alpha}_1^{(k)})$.

Step 2. Draw $\boldsymbol{\alpha}_2^{(k)}$ from $P(\boldsymbol{\alpha}_2|\mathbf{W}_2^{(k-1)}, R_2)$; draw $\boldsymbol{\theta}_2^{(k)}$ from $P(\boldsymbol{\theta}_2|\mathbf{X}_2, Y_{2,\text{obs}}, \mathbf{Y}_{-2}^{(k-1)})$; draw $Y_{2,\text{mis}}^{(k)}$ from $P(Y_2|\mathbf{X}_2, \mathbf{Y}_{-2}^{(k-1)}, \boldsymbol{\theta}_2^{(k)}, \boldsymbol{\alpha}_2^{(k)})$.

Step T. Draw $\boldsymbol{\alpha}_T^{(k)}$ from $P(\boldsymbol{\alpha}_T|\mathbf{W}_T^{(k-1)}, R_T)$; draw $\boldsymbol{\theta}_T^{(k)}$ from $P(\boldsymbol{\theta}_T|\mathbf{X}_T, Y_{T,\text{obs}}, \mathbf{Y}_{-T}^{(k-1)})$; draw $Y_{T,\text{mis}}^{(k)}$ from $P(Y_T|\mathbf{X}_T, \mathbf{Y}_{-T}^{(k-1)}, \boldsymbol{\theta}_T^{(k)}, \boldsymbol{\alpha}_T^{(k)})$.

The whole cycle is repeated, usually a small number of times, until approximate convergence. One set of MIs is then taken from the final cycle. The whole process is then repeated to produce M completed data sets. This procedure is similar to MI using chained equations (van Buuren, Brand, Groothuis-Oudshoorn, & Rubin, 2006) or sequential regressions (Raghunathan, Lepkowski, van Hoewyk, & Solenberger, 2001).

The algorithm is the possibly incompatible Gibbs sampler. Although there is no guarantee for the existence of the joint distribution from which the values are drawn,

experience has shown that it often leads to valid statistical inferences in a variety of cases (Lee & Carlin, 2010; van Buuren, 2007). The issue of compatibility of conditional distributions is a topic in ongoing and current research. Recently, Liu, Gelman, Hill, and Su (2012) studied extensively properties of the stationary distribution of the sequential imputations. They confined their attention to a subclass of incompatible models, such as our case, and showed that the estimates of parameters are consistent.

From a practical point of view, sequential imputations have great practical value since the algorithm is suitable for complex sets of incomplete data such as non-monotone missing data that do not follow a known joint distribution. The approach is widely used and has displayed success in many practical examples (Gelman & Raghunathan, 2001; Lee & Carlin, 2010; van Buuren, Boshuizen, & Knook, 1999; White, Royston, & Wood, 2011).

We now summarize the steps of the DIM as follows:

1. Impute initially missing data by taking a random draw from the observed data.
2. Repeatedly, for $t = 1,\ldots,T$:

(a). Estimate $\hat{\boldsymbol{\alpha}}_t$ in the model in equation (2) from the current completed data, and draw a random value $\dot{\alpha}_t$ from its posterior distribution.
(b). Calculate the propensity scores $\hat{\pi}_t$ given the drawn value $\dot{\alpha}_t$.
(c). Add $\hat{\pi}_t^{-1}$ into the imputation model as an additional predictor.
(d). Estimate the parameters of the imputation model $f(Y_t|\mathbf{X}, \mathbf{Y}_{-t}, \hat{\pi}_t^{-1})$ only for $Y_{t,\text{obs}}$.
(e). Draw a random value of the parameters in (d) from their posterior distributions.
(f). Impute $Y_{t,\text{mis}}$ by using the drawn value in (e) and add an appropriate amount of noise.

3. Return to step 2 to iterate the algorithm a small number of times, say 10 or 20.

When some of the estimated propensity scores are close to zero, the performance of the DIM can be disastrous. This is because some extremely large estimates of the inverse of the propensities are used for imputation of missing data. Several authors have warned against the use of inverse propensity when its variance is very large, and thus the sampling distribution of the DR imputation is skewed and highly variable (Kang & Schafer, 2007; Robins, Sued, Lei-Gomez, & Rotnitzky, 2007).

In order to evade the problem, we stratify $\hat{\pi}_t^{-1}$, $t = 1,\ldots,T$, into a few equally sized categories (usually 5). This is a classical strategy for the elimination of selection bias (Cochran, 1968). Thus, we create $s - 1$ dummy variables and include them in the imputation model as predictors. More specifically, we define dummy variables $\mathbf{d}_t = (d_{t1}, \ldots, d_{t(s-1)})$ where $d_{ts_1} = 1$ if the corresponding $\hat{\pi}_t^{-1}$ belongs to stratum s_1, $s_1 = 1, \ldots, s - 1$. The imputation model for the incomplete variable Y_t thus takes the form $f(Y_t|\mathbf{X}, \mathbf{Y}_{-t}, \mathbf{d}_t)$.

Stratification on $\hat{\pi}_t^{-1}$ can be an advantageous strategy since it mitigates the danger of $\hat{\pi}_t \approx 0$. This approach provides approximately equal weight to those individuals that are in the same stratum.

5. Simulation study

This section presents results from simulation experiments to investigate the behaviour of the DIM approach for a longitudinal study with intermittent missing data. We first assign each subject i, $i = 1,\ldots,n$, to either a control or a treatment group X_i with equal sizes and assume subjects are independent. For each visit t, $t = 1,2,3$, we generate longitudinal responses via the linear model $Y_{it} = \beta_0 + \beta_1 t + \beta_2 X_i + \beta_3 X_i t + \varepsilon_{it}$, where ε_{it} is an error component, and $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, \beta_3) = (1.0, -2.0, -1.0, 0.5)$. An unconstrained

covariance matrix Σ_y is used to generate the error components for the responses, such that the upper triangular part of the covariance matrix Σ_y is defined by $\text{vech}(\Sigma_y) = (4.0, 3.2, 2.5, 4.0, 3.0, 4.0)'$. As described in Section 3, our main interest is in making inferences about β.

In addition to the response variable $\mathbf{Y}_i = (Y_{i1}, Y_{i2}, Y_{i3})'$, we generate the secondary interest variable $\mathbf{U}_i = (U_{i1}, U_{i2}, U_{i3})'$ with another mean vector and a covariance structure such that $\text{vech}(\Sigma_u) = (6.0, 3.6, 2.0, 6.0, 3.6, 6.0)'$. Independence of the subjects implies that $\text{cov}(Y_{it}, U_{jk}) = 0$ for all t,k and $i \neq j$. But, for the same subject i, $\mathbf{Y}_i$ and $\mathbf{U}_i$ can be related. In other words, $\text{cov}(Y_{it}, U_{ik})$ is not necessarily zero for t, k = 1,2,3. We use the following structure to keep these relations simple:

$$\text{cov}(Y_{it}, U_{ik}) = \begin{cases} \sigma_{yu}, & t = k, \\ 0, & t \neq k, \end{cases}$$

for all t,k. In the simulation study, σ_{yu} took two values, 4.4 and 2.4, that produce two scenarios of strong correlation ($\rho_{yu} \approx .9$) and moderate correlation ($\rho_{yu} \approx .5$) between Y_{it} and U_{it}. It should be noted that we also considered a scenario of weak correlation ($\rho_{yu} \approx .3$), but the results were similar to the moderate scenario, so we do not report them here. We use n = 100 and n = 1,000.

To create missing data, we consider the proposed methodology by van Buuren *et al.* (2006), which is a general method for generating intermittent missing entries under MAR. We generate missing values in $\mathbf{Y}_i = (Y_{i1}, Y_{i2}, Y_{i3})'$ conditional on the observed data. For instance, for the pattern $\mathbf{R}_i = (0, 0, 1)$, missing values in Y_{i1} are created as a function of Y_{i3} and U_{i1}, and missing values in Y_{i2} are created as a function of Y_{i3} and U_{i2}. A full description of this procedure can be found in van Buuren *et al.* (2006). In our simulation, the percentages of incomplete cases for each visit were 10, 29 and 45%, respectively.

We compare the DIM approach with complete-case (CC) analysis and MI in terms of bias of the estimate of β (Bias), 95% coverage rate (Cov), and root mean squared error (RMSE). It should be noted that we calculated two types of DIM estimates: the DIM_{st} with stratification of the inverse of the propensities into five strata of equal size, and the DIM_{tr} with the inverse of the propensities censored at 0.05 when they were below 0.05. We do not report the results with the raw inverse of the propensities as the estimates went off desperately. The number of the draws of the Gibbs sampler was set to 10, and the number of imputation sets was set to 5 with 1,000 Monte Carlo simulations. All calculations were done in R 2.15.1 using MICE (version 2.13; van Buuren & Groothuis-Oudshoorn, 2011).

For MI, the imputation model is correctly specified if it consists of all observed variables $\mathbf{Y}$,$\mathbf{X}$,$\mathbf{U}$ (imp-true) or misspecified by removing the secondary interest variable $\mathbf{U}$ from the imputation model (imp-false). The complete data model can be either correct or incorrect. Likewise, for the propensity score model, there are two possibilities. Thus, for the DIM, there exist four options: both models are correctly specified ($y\otimes\pi$-true); only the propensity score model is misspecified by removing the secondary interest variable $\mathbf{U}$ from the mechanism of missingness (π-false); only the complete data model is misspecified by ignoring the fact that $\mathbf{Y}$ and $\mathbf{U}$ are related, that is, by eliminating $\mathbf{U}$ (y-false); or both models are incorrect ($y\otimes\pi$-false). It should be noted that we focus on cases where one or both models are misspecified by omitting an important variable to emphasize the importance of the secondary interest variables that are typically ignored in the imputation of missing data in practice.

Table 1. Bias, coverage (95% confidence intervals) and root mean squared error (RMSE) for the strong correlation case ($\rho_{yu} \approx .9$) with 1,000 Monte Carlo replications

		$\beta_0 = 1$			$\beta_1 = -2$			$\beta_2 = -1$			$\beta_3 = 0.5$		
		Bias	Cov	RMSE	Bias	Cov	RMSE	Bias	Cov	RMSE	Bias	Cov	RMSE
		$n = 1{,}000$											
CC		0.35	54	2.10	−0.26	18	2.23	−0.01	95	1.37	0.01	97	1.49
MI	Imp-true	0.03	97	1.85	−0.02	97	2.03	0.00	96	1.37	0.00	98	1.48
	Imp-false	0.23	80	2.01	−0.16	67	2.15	−0.01	96	1.37	0.01	97	1.49
	$y \otimes \pi$-true	0.02	97	1.84	−0.01	97	2.03	0.00	96	1.37	0.00	97	1.48
	π-false	0.03	98	1.85	−0.02	97	2.03	0.00	96	1.37	0.00	98	1.48
	y-false	0.02	98	1.85	−0.01	97	2.03	0.00	96	1.37	0.00	97	1.48
	$y \otimes \pi$-false	0.22	81	2.00	−0.15	69	2.14	−0.01	96	1.37	0.01	97	1.48
DIM_{tr}	$y \otimes \pi$-true	0.02	97	1.84	−0.01	97	2.03	0.00	96	1.37	0.00	97	1.48
	π-false	0.03	97	1.85	−0.02	98	2.03	0.00	96	1.37	0.00	98	1.48
	y-false	0.07	96	1.88	−0.05	95	2.06	0.01	96	1.37	0.00	97	1.48
	$y \otimes \pi$-false	0.22	81	2.00	−0.15	67	2.14	−0.01	96	1.37	0.01	97	1.48
		$n = 100$											
CC		0.35	92	2.16	−0.26	88	2.25	−0.01	96	1.57	0.01	97	1.53
MI	Imp-true	0.04	96	1.93	−0.03	97	2.06	−0.01	96	1.56	0.01	98	1.53
	Imp-false	0.24	94	2.08	−0.17	93	2.18	−0.01	96	1.58	0.02	97	1.54
DIM_{st}	$y \otimes \pi$-true	0.04	96	1.93	−0.03	97	2.06	−0.01	96	1.56	0.01	97	1.53
	π-false	0.05	96	1.93	−0.03	97	2.07	−0.01	96	1.57	0.01	97	1.53
	y-false	0.04	96	1.94	−0.03	97	2.06	0.00	97	1.56	0.01	97	1.53
	$y \otimes \pi$-false	0.24	96	2.08	−0.17	94	2.18	−0.01	97	1.57	0.01	97	1.54
DIM_{tr}	$y \otimes \pi$-true	0.03	97	1.92	−0.02	98	2.06	−0.01	97	1.57	0.01	98	1.53
	π-false	0.05	97	1.93	−0.03	98	2.06	−0.01	96	1.57	0.01	97	1.53
	y-false	−0.01	97	1.90	0.01	97	2.03	0.00	98	1.56	0.00	98	1.53
	$y \otimes \pi$-false	0.24	94	2.09	−0.17	94	2.18	−0.01	96	1.58	0.02	97	1.54

Note. Given are complete-case analysis (CC), standard multiple imputation (MI), and dual imputation modelling with five strata (DIM_{st}) and with truncation at 0.05 (DIM_{tr}).

Table 2. Bias, coverage (95% confidence intervals) and root mean squared error (RMSE) for the strong correlation case ($\rho_{yu} \approx .5$) with 1,000 Monte Carlo replications

		$\beta_0 = 1$			$\beta_1 = -2$			$\beta_2 = -1$			$\beta_3 = 0.5$		
		Bias	Cov	RMSE	Bias	Cov	RMSE	Bias	Cov	RMSE	Bias	Cov	RMSE
		$n = 1{,}000$											
CC		0.42	23	2.16	−0.32	00	2.28	−0.01	97	1.36	0.01	99	1.49
MI	Imp-true	0.09	95	1.89	−0.05	96	2.06	0.00	98	1.36	0.00	99	1.48
	Imp-false	0.25	66	2.02	−0.17	41	2.15	0.00	98	1.36	0.00	99	1.48
DIM_{st}	$y\otimes\pi$-true	0.08	95	1.88	−0.04	96	2.05	0.00	98	1.36	0.00	99	1.48
	π-false	0.08	95	1.89	−0.05	96	2.06	0.00	98	1.36	0.00	99	1.48
	y-false	0.08	96	1.88	−0.05	96	2.05	0.00	99	1.36	0.00	99	1.48
	$y\otimes\pi$-false	0.25	71	2.01	−0.16	45	2.15	0.00	98	1.36	0.00	99	1.48
DIM_{tr}	$y\otimes\pi$-true	0.08	94	1.89	−0.05	96	2.05	0.00	99	1.36	0.00	99	1.48
	π-false	0.09	94	1.89	−0.05	95	2.06	0.00	98	1.36	0.00	99	1.48
	y-false	0.12	93	1.92	−0.07	91	2.07	0.01	99	1.36	0.00	99	1.48
	$y\otimes\pi$-false	0.25	67	2.01	−0.16	46	2.15	−0.01	98	1.37	0.00	99	1.49
		$n = 100$											
CC		0.41	92	2.20	−0.31	84	2.28	−0.02	98	1.48	0.01	99	1.52
MI	Imp-true	0.09	98	1.95	−0.06	99	2.07	−0.02	98	1.47	0.01	99	1.51
	Imp-false	0.25	96	2.07	−0.17	96	2.16	−0.01	98	1.47	0.01	99	1.51
DIM_{st}	$y\otimes\pi$-true	0.08	98	1.94	−0.05	99	2.06	−0.02	98	1.48	0.01	99	1.51
	π-false	0.09	98	1.94	−0.06	99	2.07	−0.02	99	1.47	0.01	99	1.51
	y-false	0.09	98	1.95	−0.05	99	2.07	−0.02	98	1.48	0.01	99	1.51
	$y\otimes\pi$-false	0.24	96	2.06	−0.16	96	2.16	−0.01	99	1.48	0.01	99	1.51
DIM_{tr}	$y\otimes\pi$-true	0.07	98	1.93	−0.04	99	2.06	−0.02	99	1.48	0.01	99	1.50
	π-false	0.09	98	1.94	−0.05	99	2.07	−0.02	98	1.47	0.01	99	1.51
	y-false	0.06	98	1.93	−0.03	99	2.05	0.00	99	1.47	0.00	99	1.51
	$y\otimes\pi$-false	0.25	96	2.07	−0.16	96	2.16	−0.02	98	1.48	0.01	99	1.51

Note. Given are complete-case analysis (CC), standard multiple imputation (MI), and dual imputation modelling with five strata (DIM_{st}) and with truncation at 0.05 (DIM_{tr}).

The results for the estimates of β in the different conditions are presented in Tables 1 and 2. In general, CC performs poorly. Except for β_2 and β_3, estimates of the parameters are biased, and coverage rates are low. For instance, the coverage rate of β_1 rapidly declines to zero when the sample size is $n = 1{,}000$ in the moderate correlation situation.

As expected, MI provides correct results if the imputation model is correct (imp-true), but it produces biased estimates when the imputation model is misspecified (imp-false). The DIM strategy, on the other hand, provides valid estimates as long as one of the models is correct. Coverage rates are high too. Note that the performance of DIM_{st} is slightly better than that of DIM_{tr}, indicating a possible effect of more extreme propensities. In particular, for $n = 1{,}000$ and $\rho_{yu} \approx .5$, when the propensity score model is correctly specified only (y-false), the performance of DIM_{tr} is not as precise as that of DIM_{st}.

Interestingly, the DIM approaches perform better than CC when both models are misspecified. Also, coverage rates are improved compared to CC. Here, a misspecified imputation model is preferable to CC, but this might not be true in general. A similar situation holds for MI with incorrect imputation model (imp-false). RMSE is smaller in all situations where MI and the DIM work as expected than the other situations.

6. Application

We now demonstrate how to apply the DIM approach to the FDD introduced in Section 2. Recall that interest focuses on the effectiveness of EMDR compared to CBT. To develop the model for the data, we assume that Y_{it} represents the outcome variables PTSD-RI for child i ($i = 1,\ldots,52$) at time point t ($t = 1,2,3$) according to the longitudinal model

$$Y_{it} = \beta_0 + \beta_1 t + \beta_2 X_i + \beta_3 X_i t + \varepsilon_{it},$$

where $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, \beta_3)'$ is a vector of parameters, X_i is the treatment indicator (0 = EMDR), and $\boldsymbol{\varepsilon}_i = (\varepsilon_{i1}, \varepsilon_{i2}, \varepsilon_{i3})' \sim N(\mathbf{0}, \Sigma)$ with an unconstrained covariance matrix Σ. Had we observed full data, a classical generalized linear model could have been sufficient to obtain valid estimates of the parameters from the above model. Missing values are an inherent problem in this study, and, unfortunately, an intermittent pattern of missingness makes the analysis of FDD even more complicated. Performing an imputation strategy is a reasonable solution for estimation of β.

An important issue in FDD is relatively small sample sizes compared to the number of variables. Including all variables in the imputation model reduces the degrees of freedom for residuals that lead to overparametrization of the model. In such cases, there is not enough information in small samples compared to the number of parameters to efficiently estimate the covariance matrix. Consequently, the estimates of the parameters will become unstable. We thus should carefully construct an imputation model such that it imputes plausible values and avoids overparametrization of the model. The general advice is to include variables of scientific interest in the imputation model in addition to good predictors such as baseline covariates, and then to correct for the bias, if any, by including a function of propensities into the imputation model. This avoids enlargement of the predictors in the imputation model.

We construct the imputation model as follows. First, we include the baseline covariates sex (0 = female), age (years), treatment indicator (0 = EMDR), and indicator of the country of origin (0 = the Netherlands) into the imputation model. Second, we consider the scores for the primary outcome for the other time points. For instance, we use Y_{i1} and Y_{i3} in the imputation model for Y_{i2}. As, in addition to the outcome of interest

Table 3. Results from fireworks disaster data. Est is the parameter estimate, and SE is the standard error. CC is complete-case analysis, MI is standard multiple imputation, and DIM is dual imputation modelling

	β_0		β_1		β_2		β_3	
Method	Est	SE	Est	SE	Est	SE	Est	SE
CC	45.50	9.58	−13.81	4.67	−5.41	6.01	3.68	2.89
MI	46.76	9.91	−14.18	4.46	−7.07	6.18	3.98	2.77
DIM	47.20	12.21	−13.65	5.63	−7.29	7.57	3.89	3.47

(UCLA-RI for children), the scores from the same test were also recorded for the parents of children (U_{it}), we include these additional variables in the imputation model too. Note that to evade the growth in numbers of predictors in the imputation of Y_{it}, we only include the corresponding U_{it} in the imputation model. As an example, the imputation model for Y_{i2} contains U_{i2}, but not U_{i1} and U_{i3}. Finally, we include the propensities in the imputation model to correct for the potential bias. In Appendix A, we show a schematic representation of predictor matrix for the outcome variables Y_1, Y_2 and Y_3.

It should be mentioned that the scores for parents (U_{it}) also have missing observations. We use the same strategy for imputation. Moreover, there were further secondary outcome variables for both children and parents, but they had missing data and including them in the imputation process would lead to more intricacy. The hope is that inclusion of propensities would reduce bias that may result from misspecification of the imputation model.

To obtain propensities, we fit the propensity score model with the baseline covariates, the primary outcome variables, and the parents' PTSD-RI scores. For applying the DIM approach, we considered 10 iterations of the Gibbs sampler to produce 100 sets of imputed values. For comparison, we also estimate β by CC and MI along with the corresponding standard error. The same imputation strategy and set-up are also considered for MI.

Table 3 shows results based on the three approaches –CC, MI and DIM. A negative sign of β_1 (the coefficient of time) indicates a descending trend for both treatments EMDR and CBT over time. As can be seen from Table 3, and considering this is only one data set, the estimates of the parameters are virtually identical.

Recall the goal of the study, which is a comparison between two treatment conditions EMDR and CBT at the end of the study. The difference between two treatment conditions can be represented by $\delta = \beta_2 + 3\beta_3$. The bigger the estimated value of δ, the larger the difference between two therapies. Comparing the methods, we observe that the estimate of δ based on DIM is smaller than the other methods. That is, $\delta = 4.38$ for DIM, but $\delta = 4.87, 5.63$ for MI, CC, respectively. Although there were differences in estimates of δ, they were not statistically significant. Finally, the DIM approach produces marginally higher standard errors. The reason is that the DIM imputation model imposes more noise than standard MI.

7. Discussion

We have proposed a new imputation strategy that captures the simplicity of MI and the protection from the DR method. The proposed method introduces the concept of DR within the Markov chain Monte Carlo based framework. The DIM method as developed here can handle the problem of incomplete data in a general pattern of missingness. This

method is a useful alternative for situations where the other DR methods are difficult to apply because of the non-monotonicity of the incomplete cases.

The results from the simulation suggest that the proposed approach enhances the robustness of the imputations with respect to misspecification of the imputation model. It appears that our method avoids potential bias in estimates and inferences. Treating small propensities by stratification or truncation seems to be an effective method. Furthermore, the DIM approach is simple to implement in existing software.

In the propensity scores model we only considered the first-order terms, though a more complex model with higher-order or interaction terms could be used. Using a more complex propensity score model might theoretically improve the performance of the DIM methodology, but the pay-off in doing so is likely to be small because the subtlety in estimating the propensity scores is ignored by stratification of the propensities.

In practice, it is likely the case that both the MI model and propensity score model are misspecified to some degree. Van Buuren *et al*. (1999) distinguished between three types of variables in the imputation model: (1) variables that appear in the complete data model; (2) variables that appear in the propensity scores model; and (3) variables that explain a considerable amount of variance of the target variable. Application of DIM requires the same distinction. Current practice is to include all types in the same imputation model. The DIM model separates types 1 and 3 from type 2, with the advantage that only one of two models needs to be correct to obtain unbiased results. If both models are misspecified to some degree, DIM will not guarantee fully accurate results, as we have shown. If a good distinction is possible between types 1 and 3 versus type 2 variables, we expect that DIM is nevertheless preferable to the standard practice of throwing everything into one model. On the other hand, standard practice is likely to be easier if we cannot distinguish variable types. Finding out the relative advantages of DIM over standard practice is an interesting area of further research. In our simulation studies, the performance of the DIM approach was (marginally) better than CC when both models were wrong. It remains unknown whether these findings can be generalized to other situations.

Although we have presented the DIM approach with stratification on the inverse propensity scores, this is by no means the only approach. An alternative way to incorporate propensities into the complete data model is by utilizing a smoothing spline of the propensity scores in the complete data model (An & Little, 2008; Zhang & Little, 2009). In addition, the propensity score model in this study was misspecified by omitting important variables to emphasize the importance of the secondary interest variables. The propensity score model can also be misspecified by including or removing superfluous variables such as higher-order or interaction terms. The impact of different DIM strategies on the inferences merits further research.

Acknowledgement

The authors are grateful to the editor and two reviewers for valuable comments and suggestions which improved the paper.

References

An, H., & Little, R. J. (2008). Robust model-based inference for incomplete data via penalized splinepropensity prediction. *Communications in Statistics – Simulation and Computation*, *37*, 1718–1731. doi:10.1080/03610910802255840

Bang, H., & Robins, J. M. (2005). Doubly robust estimation in missing data and causal inference models. *Biometrics*, *61*, 962–972. doi:10.1111/j.1541-0420.2005.00377.x

Bernaards, C. A., Belin, T. R., & Schafer, J. L. (2007). Robustness of a multivariate normal approximation for imputation of incomplete binary data. *Statistics in Medicine*, *26*, 1368–1382. doi:10.1002/sim.2619

Cao, W., Tsiatis, A., & Davidian, M. (2009). Improving efficiency and robustness of the doubly robust estimator for a population mean with incomplete data. *Biometrika*, *96*, 723–734. doi:10.1093/biomet/asp033

Carpenter, J., Kenward, M. G., & Vansteelandt, S. (2006). A comparison of multiple imputation and doubly robust estimation for analyses with missing data. *Journal of the Royal Statistical Society, Series A*, *169*, 571–584. doi:10.1111/j.1467-985X.2006.00407.x

Cochran, W. G. (1968). The effectiveness of adjustment by subclassification in removing bias in observational studies. *Biometrics*, *24*, 205–213.

Daniel, R. M., & Kenward, M. G. (2012). A method for increasing the robustness of multiple imputation. *Computational Statistics and Data Analysis*, *56*, 1624–1643. doi:10.1016/j.csda.2011.10.006

Demirtas, H., Freels, S. A., & Yucel, R. M. (2008). Plausibility of multivariate normality assumption when multiply imputing non-Gaussian continuous outcomes: a simulation assessment. *Journal of Statistical Computation and Simulation*, *78*, 69–84. doi:10.1080/10629360600903866

De Roos, C., Greenwald, R., den Hollander-Gijsman, M., Noorthoorn, E., van Buuren, S., & de Jong, A. (2011). A randomised comparison of Cognitive Behavioral Therapy (CBT) and Eye Movement Desensitisation and Reprocessing (EMDR) in disaster-exposed children. *European Journal of Psychotraumatology*, *2*, 56–94. doi:10.3402/ejpt.v2i0.5694

Diggle, P. J., & Kenward, M. G. (1994). Informative dropout in longitudinal data analysis (with discussion). *Applied Statistics*, *43*, 49–93.

Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (2004). *Bayesian data analysis (*2nd ed., Chaps 6 and 7). Boca Raton, FL: Chapman and Hall/CRC.

Gelman, A., & Raghunathan, T. E. (2001). Using conditional distributions for missing data imputation. *Statistical Science*, *16*, 268–269. doi:10.1214/ss/1009213728

Jolani, S., van Buuren, S., & Frank, L. E. (2013). Combining the complete-data and nonresponse models for drawing imputations under MAR. *Journal of Statistical Computation and Simulation*, *83*(5), 866–877. doi:10.1080/00949655.2011.639773

Kang, D. Y., & Schafer, J. L. (2007). Demystifying double robustness: a comparison of alternative strategiesfor estimating a population mean from incomplete data (with discussion and rejoinder). *Statistical Science*, *22*, 523–580. doi:10.1214/07-STS227

Kenward, M. G. (1998). Selection models for repeated measurements with non-random dropout: an illustration of sensitivity. *Statistics in Medicine*, *17*, 2723–2732. doi:10.1002/(SICI1097-0258(19981215)17:23<2723::AID-SIM38>3.0.CO;2-5

Lee, K. J., & Carlin, J. B. (2010). Multiple imputation for missing data: fully conditional specification versus multivariate normal imputation. *American Journal of Epidemiology*, *171*, 624–632. doi:10.1093/aje/kwp425

Little, R. J., & Rubin, D. B. (2002). *Statistical analysis with missing data*. New York: Wiley.

Liu, J., Gelman, A., Hill, J., & Su, Y. S. (2012). On the stationary distribution of iterative imputations. Unpublished paper. Retrieved from: http://arxiv.org/abs/1012.2902

Molenberghs, G., Beunckens, C., Sotto, C., & Kenward, M. G. (2008). Every missingness not at random model has a missingness at random counterpart with equal fit. *Journal of the Royal Statistical Society, Series B*, *70*, 371–388. doi:10.1111/j.1467-9868.2007.00640.x

Qi, L., Wang, Y. F., & He, Y. (2010). A comparison of multiple imputation and fully augmented weighted estimators for Cox regression with missing covariates. *Statistics in Medicine*, *29*, 2594–2604. doi:10.1002/sim.4016

Raghunathan, T. E., Lepkowski, J. M., van Hoewyk, J., & Solenberger, P. (2001). A multivariate technique for multiply imputing missing values using a sequence of regression models. *Survey Methodology*, *27*, 85–95.

Robins, J. M., Rotnitzky, A., & Scharfstein, D. O. (1999). Sensitivity analysis for selection bias and unmeasured confounding in missing data and causal inference models. In M. E. Hallaron & D. Berry (Eds.), *Statistical models in epidemiology, the environment, and clinical trials, IMA volumes in mathematics and its applications* (pp. 1–92). New York: Springer.

Robins, J. M., Sued, M., Lei-Gomez, Q., & Rotnitzky, A. (2007). Performance of double-robust estimators when 'inverse probability' weights are highly variable. *Statistical Science*, *22*, 544–559. doi:10.1214/07-STS227D

Rosenbaum, P. R. (2002). *Observational studies*. New York: Springer-Verlag.

Rotnitzky, A. (2009). Inverse probability weighted methods. In G. Fitzmaurice, M. Davidian, G. Verbeke, & G. Molengerghs (Eds.), *Longitudinal data analysis: a handbook of modern statistical methods*(pp. 453–476). Boca Raton, FL: Chapman and Hall/CRC.

Rubin, D. B. (1976). Inference and missing data. *Biometrika*, *63*, 581–592. doi:10.1093/biomet/63.3.581

Rubin, D. B. (1987). Multiple imputation for nonresponse in surveys. New York: Wiley.

Schafer, J. L. (1997). *Analysis of incomplete multivariate data*. London, UK: Chapman and Hall.

Schafer, J. L. (2003). Multiple imputation in multivariate problems when the imputation and analysis models differ. *Statistica Neerlandica*, *57*, 19–35. doi:10.1111/1467-9574.00218

Schafer, J. L., & Graham, J. W. (2002). Missing data: our view of the state of the art. *Psycholgical Methods*, *7*, 147–177. doi:10.1037//1082-989X.7.2.147

Scharfstein, D. O., Rotnitzky, A., & Robins, J. M. (1999). Adjusting for nonignorable drop-out using semi-parametric nonresponse models (with comments). *Journal of the American Statistical Association*, *94*, 1096–1146. doi:10.1080/01621459.1999.10473862

Tsiatis, A., Davidian, M., & Cao, W. (2011). Improved doubly robust estimation when data are monotonely coarsened, with application to longitudinal studies with dropout. *Biometrics*, *67*, 536–545. doi:10.1111/j.1541-0420.2010.01476.x

van Buuren, S. (2007). Multiple imputation of discrete and continuous data by fully conditionalspecification. *Statistical Methods in Medical Research*, *16*(3), 219–242. doi:10.1177/0962280206074463

van Buuren, S. (2012). *Flexible imputation of missing data*. Boca Raton, FL: Chapman & Hall.

van Buuren, S., Boshuizen, H. C., & Knook, D. L. (1999). Multiple imputation of missing blood pressure covariates in survival analysis. *Statistics in Medicine*, *18*, 681–694. doi:10.1002/(SICI1097-0258(19990330)17:23<2723::AID-SIM71>3.0.CO;2-R

van Buuren, S., Brand, J. P. L., Groothuis-Oudshoorn, C. G. M., & Rubin, D. B. (2006). Fully conditional specification in multivariate imputation. *Journal of Statistical Computation and Simulation*, *76*, 1048–1064. doi:10.1080/10629360600810434

van Buuren, S., & Groothuis-Oudshoorn, K. (2011). MICE: multivariate imputation by chained equations in R. *Journal of Statatistical Software*, *45*, 3.

Vansteelandt, S., Carpenter, J., & Kenward, M. G. (2010). Analysis of incomplete data using inverse probability weighting and doubly robust estimators. *Methodology*, *6*, 37–48. doi:10.1027/1614-2241/a000005

White, I. R., Royston, P., & Wood, A. M. (2011). Multiple imputation using chained equations: issues and guidance for practice. *Statistics in Medicine*, *30*, 377–399. doi:10.1002/sim.4067

Zhang, G., & Little, R. J. (2009). Extensions of the penalized spline of propensity prediction method on imputation. *Biometrics*, *65*, 911–918. doi:10.1111/j.1541-0420.2008.01155.x

Zhang, G., & Little, R. J. (2011). A comparative study of doubly robust estimators of the mean with missing data. *Journal of Statistical Computation and Simulation*, *81*(12), 2039–2058. doi:10.1080/00949655.2010.516750

Received 21 March 2013; revised version received 21 June 2013

Appendix A

Referring to the FDD example, Figure 1 shows which predictor variables are used to construct imputation models for children's scores (Y_1, Y_2, Y_3). For instance, the outcome variable Y_2 is imputed using variables $Y_1, Y_3, \mathbf{X}$, and π_2^{-1} as predictors, where $\mathbf{X}$ consists of baseline covariates such as sex and treatment effect.

	Y_1	Y_2	Y_3	$\mathbf{X}$	U_1	U_2	U_3	π_1^{-1}	π_2^{-1}	π_3^{-1}
Y_1	0	1	1	1	0	0	0	1	0	0
Y_2	1	0	1	1	0	0	0	0	1	0
Y_3	1	1	0	1	0	0	0	0	0	1

Figure 1. Schematic representation of predictor matrix for the outcome variables Y_1, Y_2 and Y_3.