Dual imputation model for incomplete longitudinal data

Shahab Jolani¹*, Laurence E. Frank¹ and Stef van Buuren¹,²
¹Department of Methodology and Statistics, Utrecht University, Utrecht, The Netherlands
²Department 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

\[
E(Y_{it} | \mathbf{X}_{it}) = g(\mathbf{X}_{it}; \beta),
\]

(1)

where \( g(\mathbf{X}_{it}; \beta) \) is a known function of \( \beta \), and \( \beta \) 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 \( g(\mathbf{X}_{it}; \beta) = \beta_0 + \beta_1 t + \beta_2 A_i + \beta_3 A_i t \) with \( A_i \) defining the treatment arm indicator (CBT or EMDR), and \( \beta = (\beta_0, \beta_1, \beta_2, \beta_3)' \) 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 \( t \) 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 \( Y_{i(-t)} = (Y_{i1}, \ldots, Y_{i(t-1)}, Y_{i(t+1)}, \ldots, Y_{iT})' \), and \( \mathbf{W}_{it} = (Y_{i(-t)}, \mathbf{X}_{it}, \mathbf{U}_{it})' \). 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 \( \mathbf{a}_t \). Let \( \pi_{it}(\mathbf{a}_t) = P(R_{it} = 1 | \mathbf{W}_{it}) \); then

\[
\logit \{ \pi_{it}(\mathbf{a}_t) \} = b(\mathbf{W}_{it}; \mathbf{a}_t), \quad t = 1, \ldots, T,
\]

(2)

where \( b(\cdot; \cdot) \) is a known function. A common candidate for the transformation of \( \pi_{it}(\mathbf{a}_t) \) 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 \( \logit \{ \pi_{it}(\mathbf{a}_t) \} = a_0 + a_1 A_i + a_2 U_{it} + a_3 Y_{i(t-1)} \) 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 \( f(Y_T|X_T, Y_{-T}; \theta_T) \) to the observed part of \( Y_T \) by, for instance, least squares, and drawing \( \theta_T \) from the posterior distribution \( P(\theta_T|X_T, Y_{T,\text{obs}}, Y_{-T}) \), and (ii) imputing \( Y_{T,\text{mis}} \) by the regression estimator \( f(Y_T|X_T, Y_{-T}; \hat{\theta}_T) \) 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 \( \pi_T \). This is motivated by the fact that the propensity score is a summary of \( W_T \) that makes \( Y_T \) and \( R_T \) conditionally independent (Rosenbaum, 2002).

Following Bang and Robins (2005), we incorporate the inverse of \( \pi_T \) into the imputation model as follows. We estimate \( \pi_T \) from the parametric model in equation (2) with a known link function (e.g., logit link), and add \( \tilde{\pi}_T^{-1} \) as an additional predictor into the imputation model. Note that inclusion of \( \tilde{\pi}_T^{-1} \) is a sufficient conditions to obtain a DR estimator in this setting (Scharfstein et al., 1999). The new imputation model then fits the model \( f(Y_T|X_T, Y_{-T}, \tilde{\pi}_T^{-1}; \theta_T') = f(Y_T|X_T, Y_{-T}; \theta_T) + \gamma_T \tilde{\pi}_T^{-1} \) to the observed part of \( Y_T \) where \( \theta_T' = (\theta_T, \gamma_T) \), and \( \gamma_T \) is a regression coefficient for the new predictor \( \tilde{\pi}_T^{-1} \). As an example, the imputation model can take the linear form \( f(Y_T|X_T, Y_{-T}, \tilde{\pi}_T^{-1}; \theta_T') = \theta'_T X_T + \theta'_2 Y_{-T} + \gamma_T \tilde{\pi}_T^{-1} \), where \( \theta_T' = (\theta'_1, \theta'_2, \gamma_T) \). A random draw \( \theta_T' \) is made from its posterior distribution, and the missing part of \( Y_T \) is imputed under the regression estimator \( f(Y_T|X_T, Y_{-T}, \tilde{\pi}_T^{-1}; \theta_T') \). It should be noted that the resulting imputations are doubly protected in the sense that either the model \( f(Y_T|X_T, Y_{-T}; \theta_T) \) or the model \( \pi_T \) 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 \( Y = \{Y_{it}\} \) and \( R = \{R_{it}\} \) as \( n \times T \) matrices of outcomes and response indicators. We assume that the rows of \( Y \) are independent and identically
distributed from the joint distribution $P(Y; \theta)$ where we index the probability density function of $Y$ by $\theta$. Denote the observed part of $Y$ by $Y_{obs}$, and the missing part by $Y_{mis}$, so that $Y = (Y_{obs}, Y_{mis})$.

Under ignorability, the missingness mechanism formally plays no role in making inferences about $\theta$. 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 $V = (Y_{obs}, R, X, U)$, the posterior predictive distribution of missing data given the observed data can be defined as

$$P(Y_{mis}|V) = \int \int P(Y_{mis}|V, \theta, x)p(\theta, x|V)\,d\theta\,dx,$$  \hspace{1cm} (3)

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

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

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 $t$ 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 $x$ and estimate the propensity score $\pi_t^{-1}$ based on the drawn value. The imputation model $f(Y_t|X, Y_{-t}, \pi_t^{-1})$ then generates imputations for the missing part of $Y_t$. Note that we include $\pi_t^{-1}$ 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 $z_1^{(k)}$ from $P(z_1|W_1^{(k-1)}, R_1)$; draw $\theta_1^{(k)}$ from $P(\theta_1|X_1, Y_{1,obs}, Y_{1,obs}^{(k-1)})$; draw $Y_{1,mis}^{(k)}$ from $P(Y_1|X_1, Y_{1,obs}^{(k-1)}, \theta_1^{(k)}, z_1^{(k)})$.

**Step 2.** Draw $z_2^{(k)}$ from $P(z_2|W_2^{(k-1)}, R_2)$; draw $\theta_2^{(k)}$ from $P(\theta_2|X_2, Y_{2,obs}, Y_{2,obs}^{(k-1)})$; draw $Y_{2,mis}^{(k)}$ from $P(Y_2|X_2, Y_{2,obs}^{(k-1)}, \theta_2^{(k)}, z_2^{(k)})$.

**Step T.** Draw $z_T^{(k)}$ from $P(z_T|W_T^{(k-1)}, R_T)$; draw $\theta_T^{(k)}$ from $P(\theta_T|X_T, Y_{T,obs}, Y_{T,obs}^{(k-1)})$; draw $Y_{T,mis}^{(k)}$ from $P(Y_T|X_T, Y_{T,obs}^{(k-1)}, \theta_T^{(k)}, z_T^{(k)})$.

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 \( \hat{a}_t \) in the model in equation (2) from the current completed data, and draw a random value \( \hat{a}_t \) from its posterior distribution.
   (b) Add \( \hat{a}_t^{-1} \) into the imputation model as an additional predictor.
   (c) Add \( \hat{a}_t^{-1} \) into the imputation model as an additional predictor.
   (d) Estimate the parameters of the imputation model \( f(Y_t|X, Y_{-t}, \hat{a}_t^{-1}) \) 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 \( \hat{a}_t^{-1} \), \( t = 1, \ldots, T \), 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 \( d_i = (d_{i1}, \ldots, d_{is(s-1)}) \) where \( d_{is_1} = 1 \) if the corresponding \( \hat{a}_t^{-1} \) belongs to stratum \( s_1 \), \( s_1 = 1, \ldots, s - 1 \). The imputation model for the incomplete variable \( Y_t \) thus takes the form \( f(Y_t|X, Y_{-t}, d_i) \).

Stratification on \( \hat{a}_t^{-1} \) can be an advantageous strategy since it mitigates the danger of \( \hat{a}_t \approx 0 \). 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 \( \varepsilon_{it} \) is an error component, and \( \beta = (\beta_0, \beta_1, \beta_2, \beta_3) = (1.0, -2.0, -1.0, 0.5) \). An unconstrained
covariance matrix $\Sigma_y$ is used to generate the error components for the responses, such that the upper triangular part of the covariance matrix $\Sigma_y$ 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 $\beta$.

In addition to the response variable $Y_i = (Y_{i1}, Y_{i2}, Y_{i3})$’, we generate the secondary interest variable $U_i = (U_{i1}, U_{i2}, U_{i3})$’ with another mean vector and a covariance structure such that $\text{vech}(\Sigma_u) = (6.0, 5.6, 2.0, 6.0, 3.6, 6.0)'.$ Independence of the subjects implies that $\text{cov}(Y_{it}, U_{ij}) = 0$ for all $t,k$ and $i \neq j$. But, for the same subject $i$, $Y_i$ and $U_i$ 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, $\sigma_{yu}$ 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 $Y_i = (Y_{i1}, Y_{i2}, Y_{i3})$’ conditional on the observed data. For instance, for the pattern $R_i = (0, 0, 1)$, 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 $\beta$ (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$_{it}$ 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 $Y, X, U$ (imp-true) or misspecified by removing the secondary interest variable $U$ 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 ($\gamma \otimes \pi$-true); only the propensity score model is misspecified by removing the secondary interest variable $U$ from the mechanism of missingness ($\pi$-false); only the complete data model is misspecified by ignoring the fact that $Y$ and $U$ are related, that is, by eliminating $U$ ($\gamma$-false); or both models are incorrect ($\gamma \otimes \pi$-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 ($\rho_{\mu} \approx 0.9$) with 1,000 Monte Carlo replications

<table>
<thead>
<tr>
<th>Model Type</th>
<th>$\beta_0 = 1$</th>
<th>$\beta_1 = -2$</th>
<th>$\beta_2 = -1$</th>
<th>$\beta_3 = 0.5$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Bias</td>
<td>Cov</td>
<td>RMSE</td>
<td>Bias</td>
</tr>
<tr>
<td>$n = 1,000$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CC</td>
<td>0.35</td>
<td>54</td>
<td>2.10</td>
<td>-0.26</td>
</tr>
<tr>
<td>MI</td>
<td>0.03</td>
<td>97</td>
<td>1.85</td>
<td>-0.02</td>
</tr>
<tr>
<td>Imp-true</td>
<td>0.23</td>
<td>80</td>
<td>2.01</td>
<td>-0.16</td>
</tr>
<tr>
<td>Imp-false</td>
<td>0.02</td>
<td>97</td>
<td>1.84</td>
<td>-0.01</td>
</tr>
<tr>
<td>$\gamma \oplus \pi$-true</td>
<td>0.03</td>
<td>98</td>
<td>1.85</td>
<td>-0.02</td>
</tr>
<tr>
<td>$\pi$-false</td>
<td>0.02</td>
<td>98</td>
<td>1.85</td>
<td>-0.01</td>
</tr>
<tr>
<td>$\gamma$-false</td>
<td>0.22</td>
<td>81</td>
<td>2.00</td>
<td>-0.15</td>
</tr>
<tr>
<td>$\gamma \oplus \pi$-false</td>
<td>0.03</td>
<td>97</td>
<td>1.85</td>
<td>-0.01</td>
</tr>
<tr>
<td>DIM$_{tr}$</td>
<td>0.02</td>
<td>97</td>
<td>1.84</td>
<td>-0.01</td>
</tr>
<tr>
<td>$\gamma \oplus \pi$-true</td>
<td>0.04</td>
<td>96</td>
<td>1.93</td>
<td>-0.03</td>
</tr>
<tr>
<td>$\pi$-false</td>
<td>0.05</td>
<td>96</td>
<td>1.93</td>
<td>-0.03</td>
</tr>
<tr>
<td>$\gamma$-false</td>
<td>0.04</td>
<td>96</td>
<td>1.94</td>
<td>-0.03</td>
</tr>
<tr>
<td>$\gamma \oplus \pi$-false</td>
<td>0.24</td>
<td>96</td>
<td>2.08</td>
<td>-0.17</td>
</tr>
<tr>
<td>$n = 100$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CC</td>
<td>0.35</td>
<td>92</td>
<td>2.16</td>
<td>-0.26</td>
</tr>
<tr>
<td>MI</td>
<td>0.04</td>
<td>96</td>
<td>1.93</td>
<td>-0.03</td>
</tr>
<tr>
<td>Imp-true</td>
<td>0.24</td>
<td>94</td>
<td>2.08</td>
<td>-0.17</td>
</tr>
<tr>
<td>Imp-false</td>
<td>0.04</td>
<td>96</td>
<td>1.93</td>
<td>-0.03</td>
</tr>
<tr>
<td>$\gamma \oplus \pi$-true</td>
<td>0.05</td>
<td>96</td>
<td>1.93</td>
<td>-0.03</td>
</tr>
<tr>
<td>$\pi$-false</td>
<td>0.04</td>
<td>96</td>
<td>1.94</td>
<td>-0.03</td>
</tr>
<tr>
<td>$\gamma$-false</td>
<td>0.24</td>
<td>96</td>
<td>2.08</td>
<td>-0.17</td>
</tr>
<tr>
<td>$\gamma \oplus \pi$-false</td>
<td>0.03</td>
<td>97</td>
<td>1.92</td>
<td>-0.02</td>
</tr>
<tr>
<td>DIM$_{st}$</td>
<td>0.05</td>
<td>97</td>
<td>1.93</td>
<td>-0.03</td>
</tr>
<tr>
<td>$\gamma \oplus \pi$-true</td>
<td>-0.01</td>
<td>97</td>
<td>1.90</td>
<td>0.01</td>
</tr>
<tr>
<td>$\pi$-false</td>
<td>0.24</td>
<td>94</td>
<td>2.09</td>
<td>-0.17</td>
</tr>
</tbody>
</table>

*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

<table>
<thead>
<tr>
<th></th>
<th>$\beta_0 = 1$</th>
<th>$\beta_1 = -2$</th>
<th>$\beta_2 = -1$</th>
<th>$\beta_3 = 0.5$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Bias</td>
<td>Cov</td>
<td>RMSE</td>
<td>Bias</td>
</tr>
<tr>
<td>CC, $n = 1,000$</td>
<td>0.42</td>
<td>23.2</td>
<td>2.16</td>
<td>-0.32</td>
</tr>
<tr>
<td>MI, Imp-true</td>
<td>0.09</td>
<td>95.6</td>
<td>1.89</td>
<td>-0.05</td>
</tr>
<tr>
<td>MI, Imp-false</td>
<td>0.25</td>
<td>66.2</td>
<td>2.02</td>
<td>-0.17</td>
</tr>
<tr>
<td>DIMst, $y \otimes \pi$-true</td>
<td>0.08</td>
<td>95.6</td>
<td>1.88</td>
<td>-0.04</td>
</tr>
<tr>
<td>DIMst, $\pi$-false</td>
<td>0.08</td>
<td>95.6</td>
<td>1.89</td>
<td>-0.05</td>
</tr>
<tr>
<td>DIMst, $y$-false</td>
<td>0.08</td>
<td>96.6</td>
<td>1.88</td>
<td>-0.05</td>
</tr>
<tr>
<td>DIMst, $y \otimes \pi$-false</td>
<td>0.25</td>
<td>71.0</td>
<td>2.01</td>
<td>-0.16</td>
</tr>
<tr>
<td>DIMtr, $y \otimes \pi$-true</td>
<td>0.08</td>
<td>94.0</td>
<td>1.89</td>
<td>-0.05</td>
</tr>
<tr>
<td>DIMtr, $\pi$-false</td>
<td>0.09</td>
<td>94.0</td>
<td>1.89</td>
<td>-0.05</td>
</tr>
<tr>
<td>DIMtr, $y$-false</td>
<td>0.12</td>
<td>93.0</td>
<td>1.92</td>
<td>-0.07</td>
</tr>
<tr>
<td>DIMtr, $y \otimes \pi$-false</td>
<td>0.25</td>
<td>67.0</td>
<td>2.01</td>
<td>-0.16</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>$\beta_0 = 1$</th>
<th>$\beta_1 = -2$</th>
<th>$\beta_2 = -1$</th>
<th>$\beta_3 = 0.5$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Bias</td>
<td>Cov</td>
<td>RMSE</td>
<td>Bias</td>
</tr>
<tr>
<td>CC, $n = 100$</td>
<td>0.41</td>
<td>92.0</td>
<td>2.20</td>
<td>-0.31</td>
</tr>
<tr>
<td>MI, Imp-true</td>
<td>0.09</td>
<td>98.0</td>
<td>1.95</td>
<td>-0.06</td>
</tr>
<tr>
<td>MI, Imp-false</td>
<td>0.25</td>
<td>96.0</td>
<td>2.07</td>
<td>-0.17</td>
</tr>
<tr>
<td>DIMst, $y \otimes \pi$-true</td>
<td>0.08</td>
<td>98.0</td>
<td>1.94</td>
<td>-0.05</td>
</tr>
<tr>
<td>DIMst, $\pi$-false</td>
<td>0.09</td>
<td>98.0</td>
<td>1.94</td>
<td>-0.06</td>
</tr>
<tr>
<td>DIMst, $y$-false</td>
<td>0.09</td>
<td>98.0</td>
<td>1.95</td>
<td>-0.05</td>
</tr>
<tr>
<td>DIMst, $y \otimes \pi$-false</td>
<td>0.24</td>
<td>96.0</td>
<td>2.06</td>
<td>-0.16</td>
</tr>
<tr>
<td>DIMtr, $y \otimes \pi$-true</td>
<td>0.07</td>
<td>98.0</td>
<td>1.93</td>
<td>-0.04</td>
</tr>
<tr>
<td>DIMtr, $\pi$-false</td>
<td>0.09</td>
<td>98.0</td>
<td>1.94</td>
<td>-0.05</td>
</tr>
<tr>
<td>DIMtr, $y$-false</td>
<td>0.06</td>
<td>98.0</td>
<td>1.93</td>
<td>-0.03</td>
</tr>
<tr>
<td>DIMtr, $y \otimes \pi$-false</td>
<td>0.25</td>
<td>96.0</td>
<td>2.07</td>
<td>-0.16</td>
</tr>
</tbody>
</table>

Note. Given are complete-case analysis (CC), standard multiple imputation (MI), and dual imputation modelling with five strata (DIMst) and with truncation at 0.05 (DIMtr).
The results for the estimates of $\beta$ in the different conditions are presented in Tables 1 and 2. In general, CC performs poorly. Except for $\beta_2$ and $\beta_3$, estimates of the parameters are biased, and coverage rates are low. For instance, the coverage rate of $\beta_1$ 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_{tr} is slightly better than that of DIM_{st}, 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_{st} is not as precise as that of DIM_{tr}.

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

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

where $\beta = (\beta_0, \beta_1, \beta_2, \beta_3)'$ is a vector of parameters, $X_i$ is the treatment indicator (0 = EMDR), and $\epsilon_{it} = (\epsilon_{i1}, \epsilon_{i2}, \epsilon_{i3})' \sim N(0, \Sigma)$ with an unconstrained covariance matrix $\Sigma$. 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 $\beta$.

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
(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 \( \beta \) 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 \( \beta_1 \) (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 \( \delta \), the larger the difference between two therapies. Comparing the methods, we observe that the estimate of \( \delta \) 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 \( \delta \), 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.

### 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

<table>
<thead>
<tr>
<th>Method</th>
<th>( \beta_0 ) Est</th>
<th>( \beta_1 ) SE</th>
<th>( \beta_2 ) Est</th>
<th>( \beta_3 ) SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>CC</td>
<td>45.50</td>
<td>-13.81</td>
<td>-5.41</td>
<td>3.68</td>
</tr>
<tr>
<td>MI</td>
<td>46.76</td>
<td>-14.18</td>
<td>-7.07</td>
<td>3.98</td>
</tr>
<tr>
<td>DIM</td>
<td>47.20</td>
<td>-13.65</td>
<td>-7.29</td>
<td>3.89</td>
</tr>
</tbody>
</table>

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


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 \(Y_1, Y_3, X,\) and \(\pi_{2}^{-1}\) as predictors, where \(X\) consists of baseline covariates such as sex and treatment effect.

\[
\begin{array}{cccccccc}
Y_1 & Y_2 & Y_3 & X & U_1 & U_2 & U_3 & \pi_1^{-1} & \pi_2^{-1} & \pi_3^{-1} \\
Y_1 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\
Y_2 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\
Y_3 & 1 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \\
\end{array}
\]

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