A multiple regression imputation method with application to sensitivity analysis under intermittent missingness

Peer-reviewed author version

URANGA PINA, Rolando; MOLENBERGHS, Geert & Allende, Sira (2022) A multiple regression imputation method with application to sensitivity analysis under intermittent missingness. In: Communications in statistics. Theory and methods, 51 (15), p. 5146-5161.

DOI: 10.1080/03610926.2020.1834581
Handle: http://hdl.handle.net/1942/33320

A MULTIPLE REGRESSION IMPUTATION METHOD WITH APPLICATION TO SENSITIVITY ANALYSIS UNDER INTERMITTENT MISSINGNESS

Rolando Uranga-Piña[1], Geert Molenberghs[2] and Sira Allende-Alonso[3]

[1] Department of Data Management and Statistics, National Center for Clinical Trials, Havana, Cuba

[2] International Institute of Biostatistics and Statistical Bioinformatics, Director, Hasselt and Leuven Universities, Belgium

[3] Department of Applied Mathematics, Mathematics and Computation Building, University of Havana, Cuba

Corresponding Author: Rolando Uranga-Piña

Department of Data Management and Statistics, National Center for Clinical Trials

5th A and 60 street, Postal Code 11300, Miramar, Playa, Havana, Cuba

Email: rolando@cencec.sld.cu. Phone: (+53) 72 164 234.

Key Words: missing data; multiple imputation; sensitivity analysis; clinical trial; Gibbs sampler.

ABSTRACT

Missing data is a common problem in general applied studies, and specially in clinical trials. For implementing sensitivity analysis, several multiple imputation methods exist, like sequential imputation, which restricts to monotone missingness, and Bayesian, where the imputation and analysis models differ, entailing overestimation of variance. Also, full conditional specification provides a conditional interpretation of sensitivity parameters, requiring further calibration to get the desired marginal interpretation. We propose in this paper a multiple imputation procedure, based on a multivariate linear regression model, which keeps compatibility in sensitivity analysis under intermittent missingness, providing a marginal interpretation of the elicited parameters. Simulation studies show that the method behaves well with longitudinal data and remains robust under demanding constraints. We conclude

the possibility of situations not covered by the existing methods and well suited for our proposal, which allows more efficient handling of a given multivariate linear regression structure. Its use is illustrated in a real case study, where a sensitivity analysis is accomplished.

1. INTRODUCTION

Missing data is a common problem in general applied studies, and specially in clinical trials (Wood et al., 2004; Díaz-Ordaz et al., 2014; Uranga et al., 2017). An improper treatment of missing data may have serious implications for the accuracy of inferences of many clinical studies. Then, it is necessary to provide rigorously validated methodological tools that allow tackling this problem. Multiple imputation (MI) is a widely used method to handle missing data. It was formally introduced by Rubin (1978). Several other sources offer easy-to-read descriptions of the technique (Rubin, 1976, 2004; Buuren, 2012; Carpenter and Kenward, 2013; Little and Rubin, 2014; O'Kelly and Ratitch, 2014). It is possible to impute qualitative and quantitative data, and to distinguish among missing data in a response variable or in covariates (Verbeke and Molenberghs, 2000; Molenberghs and Verbeke, 2005).

Clinical trials typically present data of a multivariate structure where, for each individual, several response variables are measured; as well as data of a longitudinal nature, where an outcome is measured repeatedly over time. Both situations may be tackled by the so-called multivariate linear regression model. A general imputation strategy for the missing data, which is suitable in this setting, is joint modelling. Under this approach, imputations are drawn from a joint model fitted to the data. The strategy is usually implemented by a method that assumes multivariate normality, which we shall call MND (as an abbreviation of "multivariate normal data"), with a Gibbs sampler algorithm built on a somewhat restricted imputation model (Carpenter and Kenward, 2013; Buuren, 2012).

Other strategies are sequential regression imputation, which operates under monotone missingness, and full conditional specification (FCS), where the multivariate model is implicitly specified by a set of conditional univariate ones (Buuren, 2012; Carpenter and Kenward, 2013). The methods referred are currently implemented in the statistical software SAS for Windows, version 9.3 or higher (SAS, 2011). A further development is Bayesian

multiple imputation, implemented in the SAS macros by DIA Missing Data Working Group (available at the web page of the Drug Information Association missing data working group, http://missingdata.org.uk/).

Sequential MI has been used for implementing sensitivity analysis via delta-adjustment (Ratitch et al., 2013) and Bayesian MI for control-based imputation (Liu and Pang, 2017). In the first case, imputations are restricted to monotone missing data and in the second, concerns about variance overestimation have been raised (Ayele et al., 2014; Seaman et al., 2014; Liu and Pang, 2015, 2017). Tompsett et al. (2018) have criticized the conditional interpretation of sensitivity parameters under FCS and have proposed a method for improved elicitation.

Recall that Rubin's rules apply if the imputation model is more general than the analysis model, which should then be compatible, or congenial (Meng, 1994; Schafer, 1997). According to Seaman et al. (2014), the problem with control-based imputation arises because the imputer assumes more than the analyst, which is known to cause the Rubin's rule variance estimator to overestimate the repeated sampling variance (Meng, 1994). Indeed, in the so-called "jump to reference", "copy reference", and "copy increments in reference" approaches, the mean structure of the imputation model is modified, implying the imposition of strong assumptions that are no longer made when the imputed data are analysed.

Standing on concerns about complicated mean structures which could invalidate the performance of MI strategies, De Silva et al. (2017) compare, in a simulation study, multiple imputation methods for handling missing values in longitudinal data in the presence of a time-varying covariate with a non-linear association with time; De Silva et al. (2019) compare multiple imputation methods for handling missing values in a longitudinal categorical variable with restrictions on transitions over time; and Kalaycioglu et al. (2016) assess the performance of MI methods for imputing missing data in longitudinal observational health studies, with particular focus on incomplete time varying explanatory variables, while attention is also driven to covariance structures.

In this paper, a joint modelling strategy for multiple imputation of data arising from

a given multivariate linear regression model is proposed, which enhances MND by way of changing the order of drawing the parameters of the Gibbs sampler algorithm, thus allowing a more general mean structure for the imputation model. The applications of the procedure encompass data of a multivariate as well as a longitudinal nature. The method has a natural extension to account for sensitivity analysis.

Section 2 presents the methodology in six subsections. The first highlights a lack of generality inherent to the imputation model of MND. The second introduces the new proposal as a solution to this problem, with a contribution to sensitivity analysis described in Section 2.3, which constitutes a strength over sequential MI, because intermittent missing patterns are allowed; as well as over FCS, because sensitivity parameters have now a convenient marginal interpretation. To make the presentation self-contained, algebraic descriptions of the imputation algorithms of sequential MI and FCS are given. A case study is introduced in Section 2.6, and results are shown in Section 3, with two simulation studies in Sections 3.1, 3.2 and an illustrative tipping point analysis applied to the case study in Section 3.3.

2. MULTIPLE IMPUTATION METHODS AND SENSITIVITY ANALYSIS

The proposed method for multiple imputation is the main purpose of Section 2. The procedure surpasses MND in that a more general imputation model is allowed; it also surpasses sequential MI and FCS in certain applications. To this end, subsections are dedicated to the algebraic description of the imputation algorithm of each method, with one devoted to describe a contribution of the new proposal to sensitivity analysis.

2.1. *A Drawback of Multiple Imputation for Multivariate Normal Data*

We highlight in this section a drawback inherent to multiple imputation for multivariate normal data (shortly MND). Assume we have a sample from the p-variate normal distribution, denoted $\mathbf{Y}_i = (Y_{i1}, .., Y_{ip})^T, i \in \{1, ..., n\}$. The data then conform to the model

$$\mathbf{Y}_i \mid \mu, \mathbf{\Sigma} \sim \mathcal{N}(\mu, \mathbf{\Sigma}), \tag{1}$$

where μ and $\mathbf{\Sigma}$ are the mean and variance-covariance matrix of $\mathbf{Y}_i$. Assume there is partial

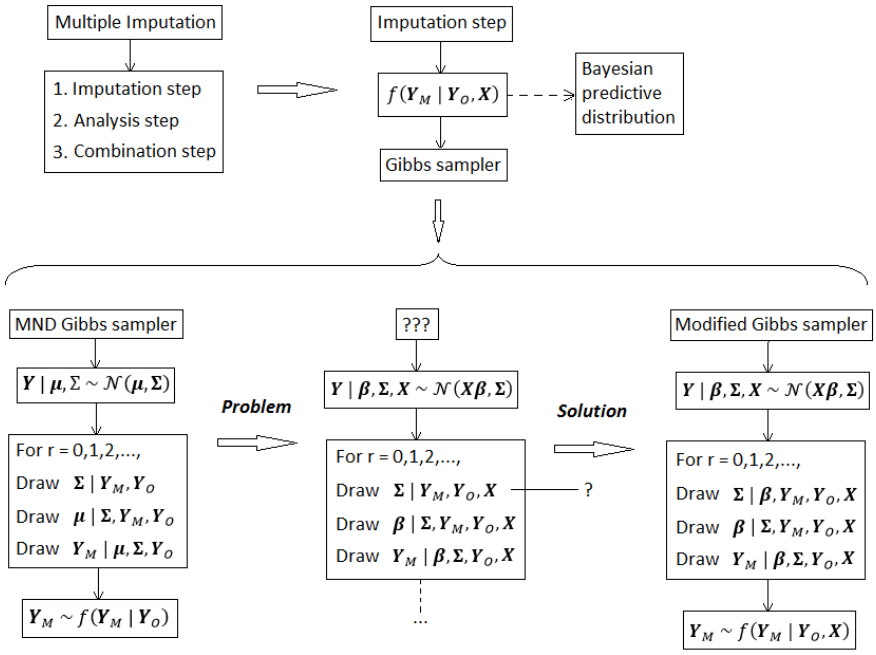

Figure 1: A drawback of MND, and a solution.

loss of information in $\mathbf{Y}$, the np column vector achieved by stacking the $\mathbf{Y}_i$, and denote by $\mathbf{Y}_O$, $\mathbf{Y}_M$ the subvectors of observed and missing data. Accept that the missingness mechanism is missing at random (MAR) (Rubin, 1976; Molenberghs and Kenward, 2007).

As Fig. 1 shows, the multiple imputation tool splits into three steps or, using Rubin's terminology, tasks: imputation, analysis and combination (Carpenter and Kenward, 2013; Rubin, 2004; Verbeke and Molenberghs, 2000). Under MAR, the imputation task entails generation of random instances from the distribution of missing given observed data, termed Bayesian predictive distribution. This distribution is usually involved, and the imputation task must be resolved by appealing to some simulation method of MCMC type (Markov Chain Monte Carlo) (Hastings, 1970) or similar, such as Gibbs sampler (Geman and Geman, 1984). To initialize the sampler, one can choose initial values of the parameters by estimation from the observed data, or draw a starting value of $\mathbf{Y}_M$ by sampling from $\mathbf{Y}_O$. In the case of MND, the Gibbs sampler switches, under an independence Jeffreys prior, between the

following steps:

1. Draw $\boldsymbol{\Sigma} \mid \mathbf{Y}_M, \mathbf{Y}_O$
2. Draw $\mu \mid \boldsymbol{\Sigma}, \mathbf{Y}_M, \mathbf{Y}_O$
3. Draw $\mathbf{Y}_M \mid \mu, \boldsymbol{\Sigma}, \mathbf{Y}_O$

to get, at the end of the chain, a random instance from $f\left(\mathbf{Y}_M \mid \mathbf{Y}_O\right)$. This is possible because in step 1 the distribution is inverse Wishart, and in steps 2 and 3 it is normal; that is, they are easy to simulate (Carpenter and Kenward, 2013; Schafer, 1997; Box and Tiao, 2011).

Assume now that, instead of (1), the data conform to the model

$$\mathbf{Y}_i \mid \beta, \boldsymbol{\Sigma}, \mathbf{X}_i \sim \mathcal{N}(\mathbf{X}_i\beta, \boldsymbol{\Sigma}) \tag{2}$$

where $\mathbf{X}_i$ are $p \times q$ matrices of known covariates. Denote by $\mathbf{X}$ the $np \times q$ matrix obtained by stacking the $\mathbf{X}_i$'s, which thus enhance the set of observed data $\mathbf{Y}_O$. We could think then about switching between

1. Draw $\boldsymbol{\Sigma} \mid \mathbf{Y}_M, \mathbf{Y}_O, \mathbf{X}$
2. Draw $\beta \mid \boldsymbol{\Sigma}, \mathbf{Y}_M, \mathbf{Y}_O, \mathbf{X}$
3. Draw $\mathbf{Y}_M \mid \beta, \boldsymbol{\Sigma}, \mathbf{Y}_O, \mathbf{X}$

to get a random instance from $f\left(\mathbf{Y}_M \mid \mathbf{Y}_O, \mathbf{X}\right)$ - the required conditional distribution of the missing given the set of observed data. This, however, is not an easy task, because we face in step 1 with an involved distribution which is no longer inverse Wishart. Hence, in this setting, the imputation and analysis models of MND would differ, causing congeniality issues (Meng, 1994). The next section introduces a multiple imputation procedure as a solution to this drawback.

2.2. *Multiple Imputation under Multivariate Linear Regression*

We have seen that the Gibbs sampler of MND does not absorb (2) as imputation model, because one of the implied conditional distributions is involved. A subtle way to overcome the problem is to attempt changing the order of drawing the parameters, as follows (see also Fig. 1):

1. Draw $\mathbf{\Sigma} \mid \beta, \mathbf{Y}_M, \mathbf{Y}_O, \mathbf{X}$
2. Draw $\beta \mid \mathbf{\Sigma}, \mathbf{Y}_M, \mathbf{Y}_O, \mathbf{X}$
3. Draw $\mathbf{Y}_M \mid \beta, \mathbf{\Sigma}, \mathbf{Y}_O, \mathbf{X}$

This happens to be a satisfactory solution: under an independence Jeffreys prior, the distribution in step 1 is inverse Wishart, and in steps 2 and 3 it is normal. We shall call the presented proposal to MI: MLR, as an abbreviation of "multivariate linear regression". To obtain the explicit expressions for the conditional distributions included in the Gibbs sampler algorithm of MLR choose, following Box and Tiao (2011), the non-informative prior

$$f(\beta, \mathbf{\Sigma}) \propto |\mathbf{\Sigma}|^{-\frac{p+1}{2}}.$$

Then the posteriors implied by model (2) are

$$f(\mathbf{\Sigma} \mid \beta, \mathbf{Y}, \mathbf{X}) \equiv W^{-1}\left(n, \sum_{i=1}^{n} (\mathbf{Y}_i - \mathbf{X}_i\beta)(\mathbf{Y}_i - \mathbf{X}_i\beta)^T\right),$$

$$f(\beta \mid \mathbf{\Sigma}, \mathbf{Y}, \mathbf{X}) \equiv \mathcal{N}\left(\left(\mathbf{X}^T(\mathbf{I}_n \otimes \mathbf{\Sigma}^{-1})\mathbf{X}\right)^{-1}\mathbf{X}^T(\mathbf{I}_n \otimes \mathbf{\Sigma}^{-1})\mathbf{Y}, \left(\mathbf{X}^T(\mathbf{I}_n \otimes \mathbf{\Sigma}^{-1})\mathbf{X}\right)^{-1}\right)$$

where $\mathbf{I}_n$ is the identity matrix of order n. Note that these expressions are the same as formulas (7) and (8) of Percy (1992). To get $f(\mathbf{Y}_M \mid \beta, \mathbf{\Sigma}, \mathbf{Y}_O, \mathbf{X})$, denote by $\mathbf{X}_O$, $\mathbf{X}_M$ the submatrices of $\mathbf{X}$ corresponding to $\mathbf{Y}_O$, $\mathbf{Y}_M$, by

$$\begin{pmatrix} \mathbf{\Sigma}_{OO} & \mathbf{\Sigma}_{OM} \\ \mathbf{\Sigma}_{MO} & \mathbf{\Sigma}_{MM} \end{pmatrix}$$

the resultant grouping of the matrix $\mathbf{I}_n \otimes \mathbf{\Sigma}$, and apply the formula of the multivariate conditional normal distribution to

$$\begin{pmatrix} \mathbf{Y}_O \\ \mathbf{Y}_M \end{pmatrix} \mid \beta, \mathbf{\Sigma}, \mathbf{X} \sim \mathcal{N}\left(\begin{pmatrix} \mathbf{X}_O\beta \\ \mathbf{X}_M\beta \end{pmatrix}, \begin{pmatrix} \mathbf{\Sigma}_{OO} & \mathbf{\Sigma}_{OM} \\ \mathbf{\Sigma}_{MO} & \mathbf{\Sigma}_{MM} \end{pmatrix}\right)$$

yielding

$$f\left(\mathbf{Y}_M \mid \beta, \mathbf{\Sigma}, \mathbf{Y}_O, \mathbf{X}\right) \equiv \mathcal{N}(\mathbf{X}_M\beta + \mathbf{\Sigma}_{MO}\mathbf{\Sigma}_{OO}^{-1}\left(\mathbf{Y}_O - \mathbf{X}_O\beta\right), \mathbf{\Sigma}_{MM} - \mathbf{\Sigma}_{MO}\mathbf{\Sigma}_{OO}^{-1}\mathbf{\Sigma}_{OM}).$$

We shall see next that not only MLR surpasses MND in a greater flexibility for the imputation task, but also enjoys a natural application to sensitivity analysis, namely to the so-called delta-adjustment where, in contrast to previous methods proposed in the statistical literature, the analysis can be done under general intermittent patterns of missing data, and the sensitivity parameters have a convenient marginal interpretation.

2.3. *Application of MLR to sensitivity analysis*

Delta-adjusted pattern imputation is a strategy that can be used for sensitivity analysis under a clearly formulated clinical assumption (Ratitch et al., 2013; Liu and Pang, 2017). Specifically, the assumption is that subjects from the experimental treatment arm in a controlled clinical trial who miss a given time-point would have, on average, their unobserved efficacy score worse by some amount δ compared with the observed efficacy score of subjects that continue to the next time-point. Subjects missed from the control arm would exhibit the same evolution of the disease as control subjects that stay on study. Ratitch et al. (2013) restrict to monotone missingness and Liu and Pang (2017) overcome intermittent missingness patterns by assuming that the intermittent missing data can be imputed under MAR to create monotone patterns.

MLR can be successfully used for applying delta-adjustment in the presence of intermittent missing data. To this end assume, in the context of Sections 2.1 and 2.2, that instead of (2), the data conform to the model

$$\mathbf{Y}_i \mid \beta, \delta, \mathbf{\Sigma}, \mathbf{X}_i, \mathbf{Z}_i \sim \mathcal{N}(\mathbf{X}_i\beta + \mathbf{Z}_i\delta, \mathbf{\Sigma}) \tag{3}$$

where $\mathbf{Z}_i$ are additional covariates and δ is a set of sensitivity parameters whose values would be elicited (Tompsett et al., 2018). Then, after imputing missing values of $\mathbf{Y}_i$ using (2) as imputation model via MLR, simply add the additional constant term $\mathbf{Z}_i\delta$ to get the desired imputed data from model (3).

This procedure corresponds to the approach, described in Ratitch et al. (2013), that

first imputes all values (all time-points) using an MAR-based method and only then applies δ-adjustments. Typically, $\mathbf{Z}_i$ would be a $p \times 1$ vector of missing data indicators, weighted by some constants (which vanish for the control group in a clinical trial), and δ a scalar quantity. According to Ratitch et al. (2013), the procedure can be useful if it is desired to impose a fixed and definite set of quantities to encapsulate the change in efficacy associated with missingness - for example, to impose a worsening over time that was observed in withdrawals in historic data.

The proposed approach to sensitivity analysis via MLR surpasses previous approaches based on sequential MI, which restrict to monotone missing data, and FCS, where the sensitivity parameters have a conditional interpretation. To make the exposition self-contained, we give an algebraic description of the imputation algorithms of these last two methods.

2.4. *Sequential Regression Imputation*

Sequential regression is a method of multiple imputation described in Carpenter and Kenward (2013). It works under MAR monotone missingness. With notations borrowed from Section 2.1, the method assumes, at each stage j, $2 \leq j \leq p$, the validity of the univariate linear regression model

$$Y_{ij} = \gamma_{0j} + \gamma_{1j} Y_{i1} + \ldots + \gamma_{j-1,j} Y_{i,j-1} + \varepsilon_{ij}, \varepsilon_{ij} \sim \mathcal{N}\left(0, \sigma_j^2\right).$$

Using data from individuals with Y_{ij} observed which, by the monotone assumption, have $Y_{i1}, \ldots, Y_{i,j-1}$ observed, fit this model, obtaining the ordinary least squares estimates $\widehat{\gamma}_j, \widehat{\sigma}_j^2$ of $\gamma_j = (\gamma_{0j}, \gamma_{1j}, \ldots, \gamma_{j-1,j})$ and σ_j^2. Assuming that the prior distribution of $\left(\gamma_j, \sigma_j^2\right)$ has density proportional to σ_j^{-2}, sequential imputation performs the following two steps in the imputation task:

1. Draw $\gamma_j, \sigma_j^2 \mid \widehat{\gamma}_j, \widehat{\sigma}_j^2$
2. For each unobserved Y_{ij}, draw $Y_{ij} \mid Y_{i1}, \ldots, Y_{i,j-1}, \gamma_j, \sigma_j^2$

This is possible because the distribution in step 1 is a particular case of the normal-inverted Wishart, and in step 2 it is normal. Note that for $j = 3, \ldots, p$ there will be

some units with Y_{ij} missing and with one or more of $Y_{i2}, \ldots, Y_{i,j-1}$ missing, and imputed at previous steps. These previously imputed values are used when imputing Y_{ij}. It is assumed that Y_{i1} is observed, $1 \leq i \leq n$. If appropriate in the context, the fully observed variables can include covariates, like group indicators in a randomized controlled study. To allow for intermittent missing data, one can first use MND to transform intermittent into monotone, then use sequential MI. Ratitch et al. (2013) report an application of Sequential MI to sensitivity analysis under monotone missing data.

2.5. *Full Conditional Specification*

Full conditional specification is described in Buuren (2012). With a slight change of notation from Section 2.1, let the data be represented by the $n \times p$ matrix $\mathbf{Y}$. In the presence of missing data $\mathbf{Y}$ is partially observed. Denote by $\mathbf{Y}_j$ the j-th column in $\mathbf{Y}$, and by $\mathbf{Y}_{-j}$ the complement of $\mathbf{Y}_j$, that is, all columns in $\mathbf{Y}$ except $\mathbf{Y}_j$. Under MAR, the algorithm for the imputation task of FCS is as follows.

1. Specify an imputation model $P\left(\mathbf{Y}_j^{miss} \mid \mathbf{Y}_j^{obs}, \mathbf{Y}_{-j}, \theta_j\right)$ for variable $\mathbf{Y}_j$ with $j = 1, \ldots, p$, where θ_j are parameters.
2. For each j, fill in starting imputations $\widetilde{\mathbf{Y}}_j^0$ by random draws from $\mathbf{Y}_j^{obs}$.
3. Repeat for $t = 1, \ldots, T$:
4. Repeat for $j = 1, \ldots, p$:
5. Define $\widetilde{\mathbf{Y}}_{-j}^t = \left(\widetilde{\mathbf{Y}}_1^t, \ldots, \widetilde{\mathbf{Y}}_{j-1}^t, \widetilde{\mathbf{Y}}_{j+1}^{t-1}, \ldots, \widetilde{\mathbf{Y}}_p^{t-1}\right)$ as the currently complete data except $\mathbf{Y}_j$.
6. Draw $\widetilde{\theta}_j^t \sim P\left(\theta_j^t \mid \mathbf{Y}_j^{obs}, \widetilde{\mathbf{Y}}_{-j}^t\right)$.
7. Draw imputations $\widetilde{\mathbf{Y}}_j^t \sim P\left(\mathbf{Y}_j^{miss} \mid \mathbf{Y}_j^{obs}, \widetilde{\mathbf{Y}}_{-j}^t, \widetilde{\theta}_j^t\right)$.
8. End repeat j.
9. End repeat t.

Tompsett et al. (2018) criticize the use of FCS in sensitivity analysis, due to the conditional interpretation of parameters, and develop procedures for tackling this problem. In this respect note that, with the extension of MLR to account for delta-adjustment, the sensitivity parameters have a marginal meaning, which is a strength.

Table 1: Missingness patterns for the Hamilton Depression study.

Measurement occasion								
Week 1	Week 2	Week 3	Week 4	Week 6	Week 8	Week 10	Number	%
Completers								
O	O	O	O	O	O	O	52	67.53
Dropouts								
O	O	O	O	O	O	M	6	7.79
O	O	O	O	O	M	M	2	2.60
O	O	O	M	M	M	M	2	2.60
O	O	M	M	M	M	M	4	5.19
O	M	M	M	M	M	M	5	6.49
M	M	M	M	M	M	M	2	2.60
Non-monotone missingness								
O	O	M	O	O	O	O	2	2.60
O	M	O	O	M	M	M	1	1.30
O	M	O	M	M	M	M	1	1.30

2.6. *Case Study*

The sensitivity analysis described in this paper was applied to data from a phase II, multicenter, randomized, double-blind, placebo-controlled clinical trial in subjects with major depressive disorder. The primary endpoint was the total score from the first 17 questions of the 25-item Hamilton Depression Scale (HAM-D-17 total score) ranging from 0 to 52, with lower values corresponding to more favourable outcomes. Study visits were scheduled at baseline and at weeks 1, 2, 3, 4, 6, 8, and 10. The aim of the trial was to assess if the arms differed in the HAM-D-17 total score at week 10, after adjusting for baseline values. Table 1 shows the missingness patterns: 21/77 patients (27.3%) have monotone missing data and 4 subjects (5.2%) have intermittent missingness.

We used the "tipping point" strategy for analysis of this incomplete data set, where the delta-adjusting approach discussed in Section 2.3 was used for estimating sensitivity to a range of assumptions (Ratitch et al., 2013). A series of analyses were performed with a range of increasing δ values so that we could assess at which point the study conclusions changed from favourable to unfavourable, that is, so that we could find a tipping point. Then, a clinical interpretation regarding the δ value representing the tipping point was considered to judge whether the corresponding differences between the dropouts and completers were plausible. To define $\mathbf{Z}_i$ we assigned, to each subject i in the experimental arm, increasing weights of consecutive values per each run of missing values.

The following model was used for the analysis task:

$$\begin{cases} Y_{ij} = \beta_1 + \beta_2 H_i + \beta_3 G_i + \beta_4 t_j + \beta_5 G_i t_j + \beta_6 t_j^2 + \beta_7 z_{ij} + b_i + \varepsilon_{ij}, \\ b_i \sim \mathcal{N}\left(0, d^2\right), \varepsilon_{ij} \sim \mathcal{N}\left(0, \sigma^2\right) \ \textit{independent}. \end{cases} \tag{4}$$

Here $1 \leq i \leq 77$, $1 \leq j \leq 7$; H_i is the baseline HAM-D-17 total score of patient i; G_i takes the values 0 for control and 1 for the experimental group; the time variable t_j represents measurement occasion in weeks; and z_{ij} is the weighted missing data indicator. To impute missing data via MLR under the MAR assumption, we used a model with the same mean structure but the term "$\beta_7 z_{ij}$" excluded, and with an unstructured covariance matrix. We then added the constant term "δz_{ij}" to perform delta-adjustment. Under MAR, the effect of interest is given by $\beta_3 + 10\beta_5$.

3. RESULTS

To provide an experimental validation of MLR, two simulation studies are undertaken, the first in a standard longitudinal setting and the second, in a more demanding setting, where we shall see that MLR surpasses its counterparts. The exposition ends with results from a sensitivity analysis performed to the case study described in Section 2.6.

3.1. *Simulation Study in a Standard Setting*

We present in this section the results of a simulation study in a standard longitudinal setting; the aim is to compare the performance of MLR with its analogues implemented in

SAS. Assume the execution of a hypothetical clinical trial with $N = N_0 + N_1$ individuals enrolled in a control (N_0 subjects) and an experimental (N_1 subjects) group. For each patient one plans to collect measurements of a quantitative response variable at p occasions. The profile of a hypothetical individual i is defined as the vector of p components $\mathbf{Y}_i = (Y_{i1}, ..., Y_{ip})$ where each component Y_{ij} represents the measurement of the response variable at occasion j. We use the following instance of (2) as analysis model:

$$\begin{cases} Y_{ij} = \beta_1 + \beta_2 G_i + \beta_3 t_j + \beta_4 G_i t_j + \delta_{ij}, \\ \delta_{ij} = \alpha \delta_{i,j-1} + z_{ij}, \delta_{ij} \sim \mathcal{N}\left(0, \lambda^2\right), \\ z_{ij} \perp\!\!\!\perp \delta_{i1}, ..., \delta_{i,j-1}, \end{cases} \tag{5}$$

allowing an unstructured covariance matrix for the imputation model.

In these notations, G_i takes the values 0 for control and 1 for the experimental group; the time variable t_j takes the value $\frac{j}{p}$, $1 \leq j \leq p$, j integer; the parameter vector is $\theta = (\beta_1, \beta_2, \beta_3, \beta_4, \alpha, \lambda^2)$. We set $N_0 = N_1 = 75$ and generate random samples from model (5) with $p = 3$, $\theta = (9, 4, 8, 3, 0.6, 2)$; then generate missing data according to the logistic model

$$logit\left[P\left(D_i = j \mid D_i \geq j, \mathbf{y}_i\right)\right] = \psi_0 + \psi_1 y_{ij} + \psi_2 y_{i,j-1},\ 2 \leq j \leq p,$$

in line with Diggle and Kenward (1994), setting $\psi_0 = 0, \psi_1 = 0, \psi_2 = -0.085$. This is an MAR mechanism restricted to dropout, where D_i denotes dropout occasion. At the measurement level, the proportions of missing values are equal to approximately 20%.

We use one of the following methods to get a maximum likelihood (ML) estimate $\widehat{\theta}$ of θ with an associated 95% confidence interval (CI) for each component: direct ML, MND, sequential MI, FCS, and MLR. The process is iterated 500 times and allows to get, as summary indicators of the performance of the methods, coverage, defined as the proportion with which the CI covers the true value, and bias, defined as the absolute value of the average difference between true value and estimate. Results are shown in Table 2, where it is apparent that MLR and its counterparts behave excellently, with negligible bias and coverage about 95%. SAS code is given in Part I of an electronic appendix.

3.2. *Simulation Study in a Demanding Setting*

Table 2: Absolute bias and coverage of five methods for treating missing data in a standard longitudinal setting (ML = maximum likelihood, MND = multiple imputation for multivariate normal data, Sequential MI = multiple imputation by sequential regression, FCS = full conditional specification, MLR = multivariate multiple imputation).

	Setting									
	ML		MND[1]		Sequential MI		FCS		MLR	
	$\beta_1 = 9,\ \beta_2 = 4,\ \beta_3 = 8,\ \beta_4 = 3,\ \alpha = 0.6,\ \lambda^2 = 2$									
	Bias	Cov.	Bias	Cov.	Bias	Cov.	Bias	Cov.	Bias	Cov.
β_1	0.00	0.94	0.00	0.95	0.00	0.94	0.00	0.93	0.00	0.94
β_2	0.00	0.93	0.00	0.94	0.01	0.94	0.00	0.93	0.00	0.93
β_3	0.01	0.93	0.00	0.94	0.00	0.96	0.01	0.94	0.00	0.96
β_4	0.00	0.94	0.01	0.93	0.01	0.95	0.00	0.94	0.00	0.93
α	0.00	0.96	0.00	0.96	0.00	0.96	0.00	0.96	0.00	0.95
λ^2	0.01	0.95	0.03	0.96	0.01	0.94	0.02	0.95	0.01	0.95

[1] Multiple imputations are performed separately by each category of the group covariate.

Here we present the results from a second simulation study. Similar to Section 3.1, assume the execution of a hypothetical clinical trial, but now label the group indicator as $G_i = 1$ for the control and $G_i = 0$ for the experimental arm. We then use the model

$$\begin{cases} Y_{ij} = \beta_1 + \beta_2 G_i + \beta_3 t_j + \delta_{ij}, 1 \leq j \leq 2, \\ \delta_{i2} = \alpha\delta_{i1} + z_{i2}, \delta_{ij} \sim \mathcal{N}\left(0, \lambda^2\right), \\ z_{i2} \perp\!\!\!\perp \delta_{i1} \end{cases} \tag{6}$$

and set $N_0 = N_1 = 75$, $p = 2$, $\theta = (\beta_1, \beta_2, \beta_3, \alpha, \lambda^2) = (9, 4, 8, 0.6, 2)$. Following generation of random samples from model (6), values Y_{ij} are deleted with $G_i = 1,\ j = 2$. This is an instance of an MAR mechanism, dependent on the group covariate G_i. The number of iterations is 500. We assess the performance of the same methods described in Section 3.1. Results are shown in Table 3.

Table 3: Average estimates, bias and coverage of five methods of treating missing data in a demanding setting (ML = maximum likelihood, MND = multiple imputation for multivariate normal data, Sequential MI = multiple imputation by sequential regression, FCS = full conditional specification, MLR = multivariate multiple imputation).

	Setting								
	ML			MND[1]	Sequential MI	FCS	MLR		
	$\beta_1 = 9,\ \beta_2 = 4,\ \beta_3 = 8,\ \alpha = 0.6,\ \lambda^2 = 2$								
	Est.	Bias	Cov.				Est.	Bias	Cov.
β_1	8.99	0.01	0.95				9.00	0.00	0.95
β_2	4.00	0.00	0.94	Algorithm	Algorithm	Algorithm	4.00	0.00	0.94
β_3	8.01	0.01	0.96	fails	fails	fails	8.00	0.00	0.94
α	0.60	0.00	0.95				0.60	0.00	0.94
λ^2	1.98	0.02	0.95				2.00	0.00	0.96

[1] Multiple imputations are performed separately by each category of the group covariate.

Sequential MI fails because, to impute, it works with the conditional distribution of the second measurement given the first. Since, for the observed second measurements Y_{i2}, the group indicator G_i is identically zero, the group effect β_2 does not appear in the imputation algorithm, and this implies that β_2 is inestimable with this method. FCS uses, in turn, the two conditional distributions implied by model (6) for Y_{i1} given Y_{i2} and vice-versa. Here occurs, again, the phenomenon encountered with Sequential MI, when imputing the second observation given the first.

MND fails because imputations are being done separately for each category of the group covariate, and for $G_i = 1$ there are not observations on Y_{i2}, which causes non-convergence of the algorithm. This phenomenon resembles a similar one with maximum likelihood: the ML estimate of μ in model (1) does not exist if, for all i and some fixed j, the j-th component of $\mathbf{Y}_i$ is missing. One could attempt fitting the joint model differently, e. g. altogether.

However, we should not expect good behaviour in this case, since the imputation model does not include categorical predictors, implying that the analysis model (6) cannot be embedded in it (the models are not compatible). An additional simulation was done and reported distorted results (high bias and low coverage).

The explanation why MLR behaves well in this setting is the same as that for direct ML: these methods work in the long format, allowing for categorical predictors. Despite the missing data structure, the maximum likelihood estimator from model (6) exists, and imputations via MLR are also well-defined.

3.3. *Tipping Point Analysis for the Hamilton Depression Clinical Trial*

Table 4 and Fig. 2 present the results of analyses using the MLR delta-adjustment strategy with different values of δ applied to the example dataset described in Section 2.6. The values of δ range from 0 (corresponding to an MAR based MI analysis) to 3. These results illustrate how conclusions for early missed observations could be affected by increasingly pessimistic assumptions for the experimental arm about the HAM-D-17 score.

Under MAR, the estimated profile for non-completers coincides, in both groups, with that of completers, because the coefficient of the missingness indicator is set to zero. Since, for patients in the control group, completers and non-completers are assumed to share the same behaviour, their predicted profiles also coincide in all settings. When $\delta = 0$, the predicted profile in the experimental arm is significantly separated, in favour of a better response, from that of the control arm. When the HAM-D-17 is increased by $\delta = 1$ at each consecutive missed visit, the estimated profile for a patient who drops out at the fourth week in the experimental arm is closer to that of the controls, as shown in Fig. 2. When $\delta = 2$, experimental dropouts surpass controls in favour of a worse response, and when $\delta = 3$, this inverted separation is more pronounced.

Table 4 shows the estimates of the group effect at week 10 under the different values of δ. When $\delta = 0, 1, 2$, a significant effect in favour of the experimental group is attained. When $\delta = 3$, significance is lost. If a worsening of 3 in the HAM-D-17 score is clinically plausible for early non-completers in the experimental arm, then this tipping point analysis would

Table 4: Results from analyses using standard multiple imputation (MI) and delta-adjustment. Estimates of group effect at week 10 of Hamilton Depression Scale total score.

Analysis	Estimate (CI)	p-value
Standard MI (MAR-based)	-6.503 (-8.980,-4.027)	0.0000
Delta-adjusting: $\delta = 1$	-4.640 (-6.968,-2.312)	0.0001
Delta-adjusting: $\delta = 2$	-3.346 (-6.012,-0.680)	0.0143
Delta-adjusting: $\delta = 3$	-2.396 (-5.235, 1.043)	0.1894
CI, 95% confidence interval.		

suggest that the MAR result of the study needs to be treated with caution. As suggested by Ratitch et al. (2013), a change of approximately one-half of a standard deviation is clinically meaningful for a patient. In our example data set the mean of HAM-D-17 total score at baseline was 23.1, and standard deviation was 2.56 for the two treatment groups combined. Therefore, a $\delta = 2$ would be a meaningfully pessimistic assumption, yet study conclusions would still hold, whereas $\delta = 3$ might be considered as quite conservative. SAS code is given in Part III of an electronic appendix.

4. CONCLUSIONS AND DISCUSSION

The Guideline on missing data in confirmatory clinical trials stresses that "A positive regulatory decision must be based on an analysis where the possibility of important bias in favour of the experimental agent can be excluded" (EMA, 2010). This document recommends using likelihood-based methods as the primary analysis, and conducting sensitivity analysis to assess the robustness of the results under certain deviations from the MAR assumption (EMA, 2010; NRC, 2010; Liu and Pang, 2017). According to Carpenter and Kenward (2013), "Given a set of assumptions about the reasons data are missing, there are a number of statistical methods for carrying out the analysis. Nevertheless, we argue that none shares the practical utility, broad applicability and relative simplicity of Rubin's Multiple Imputation."

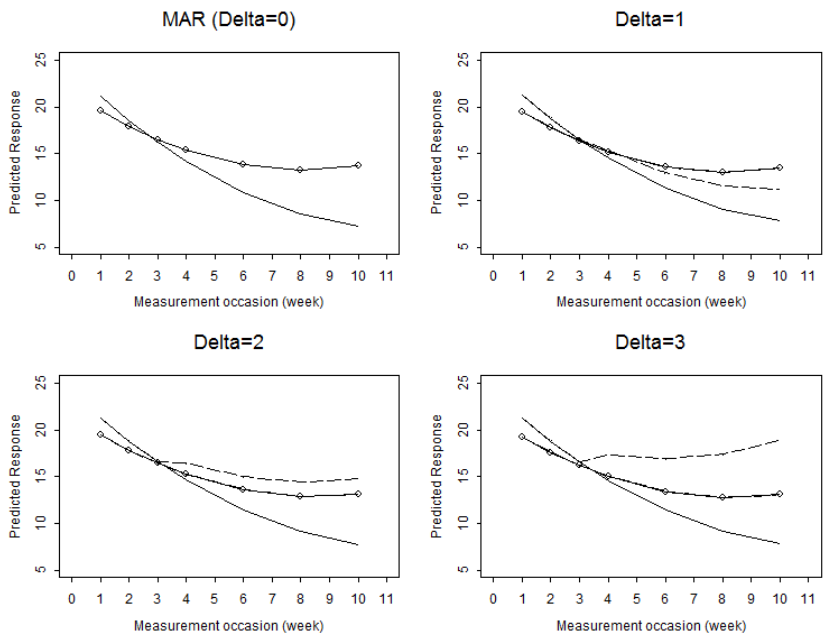

Figure 2: Tipping point analysis for the Hamilton Depression study. Solid lines: completers; dash lines: dropouts at fourth week; dots: experimental group; circles: control group.

We introduced in this paper a joint modelling strategy for multiple imputation of data consistent with a given multivariate linear regression model, which we named MLR. The procedure surpasses MND in that a more general mean structure for the imputation model is allowed, attained by way of changing the order of drawing the parameters of the Gibbs sampler algorithm. Specifically, while the imputation model of MND does not distinguish covariates explicitly, that of MLR does.

A simulation study was performed in a standard longitudinal setting, where the proposed method and several existing procedures behaved equally well. We also identified a scenario, called "demanding setting", where the standard methods sequential MI and FCS fail, due to their wide format imputation algorithms, whereas the new proposal works, because it imputes in the long format. Although MND imputes in the long format too, it fails because its imputation model does not include categorical predictors, implying the need of performing imputations separately by each category. The situation can be found in clinical trials on

strongly impaired patients, for example some sort of cancer, where it is not possible to collect measurements at the second occasion for patients in the control group.

An application of MLR to sensitivity analysis was revealed; namely, to the so-called delta-adjustment setting, where the assumption is that subjects from the experimental treatment arm in a controlled clinical trial who miss a given time-point would have, on average, their unobserved efficacy score worse by some amount δ compared with the observed efficacy score of subjects that continue to the next time-point. MLR can be successfully applied to this setting, whether the patterns of missing data are monotone or intermittent. This could be considered an improvement over previous works that restrict to monotone missingness (Ratitch et al., 2013; Liu and Pang, 2017).

Control-based imputation is an approach to sensitivity analysis that imputes missing data in the test drug group using a model built from the control group. In this setting, concerns about variance estimation have been raised, because the imputation models are different from the final analysis model, causing congeniality issues (Ayele et al., 2014; Seaman et al., 2014; Liu and Pang, 2015, 2017; Meng, 1994; Nielsen, 2003). It has been pointed out that the mean structure of the imputation model, bearing in procedures like "jump to reference", "copy reference", and "copy increments in reference", entails more assumptions of the imputer than the analyst, which is known to cause the Rubin's rule variance estimator to overestimate the repeated sampling variance (Seaman et al., 2014; Meng, 1994). With MLR, this should not be anymore the case, because the mean structure of the imputation and analysis models can be equated in a natural way.

Owing to concerns about complicated mean structures, which could invalidate the performance of MI strategies, De Silva et al. (2017) performed a comparison of multiple imputation methods for handling missing values in longitudinal data in the presence of a time-varying covariate with a non-linear association with time. The methods included FCS, MND, and two-fold FCS. Once again, the difference between the imputation and analysis models was the motivating reason of investigation, because of the danger of uncongeniality issues. MLR could be included in these types of simulations, taking into account that it naturally incor-

porates non-linear predictors, due to the conditioning on covariates.

Tompsett et al. (2018) propose a variant of FCS for elicitation of sensitivity parameters via calibration. They consider models in which the sensitivity parameters have a conditional interpretation, which is not practical. Then they match those models with others built in such a way that the sensitivity parameters have a convenient marginal interpretation. The matching is algebraically and computationally involved. In this respect we have noted that, with the extension of MLR to account for delta-adjustment, the sensitivity parameters have a marginal meaning, which is a strength.

Specifically, following Tompsett et al. (2018), in a not at random fully conditional specification (NARFCS) procedure, the sensitivity parameters are the specified differences between imputed and observed values of a variable, conditional on all remaining variables of the data and their missingness indicators; hence they are termed conditional sensitivity parameters, or CSP. When a sensitivity parameter is marginal on at least some of the remaining variables, it is referred to as marginal sensitivity parameter (MSP). In contrast to MSP, direct elicitation of a CSP is typically not feasible, as this conditional nature forces one to elicit information about groups of people who are matched in ways that are not commonly analysed, if at all. This makes typical elicitation nearly impossible for NARFCS, and the authors provide an algorithm to perform calibration, by relating the CPS to MSP. In the case of MLR, the sensitivity parameters are marginal on the remaining variables, albeit conditional on the missingness indicators; hence they are MSP, and easier to interpret.

As a limitation of MLR, we should stress that, although the mean structure of the multivariate linear regression model (2) is fully absorbed in a natural fashion, the covariance matrix $\boldsymbol{\Sigma}$ remains unstructured, generating $\frac{p(p+1)}{2}$ variance-covariance parameters. This does not affect validity of inferences if the analysis model has a simpler structured matrix (for example compound-symmetric or first-order autoregressive) because, in this case, the imputation model is more general than the analysis model (Meng, 1994; Schafer, 1997). As a coming work, we intend to extend the method to include missing data in covariates, which could allow enhancing the simulation studies of Kalaycioglu et al. (2016).

ACKNOWLEDGEMENTS

Financial support from the VLIR-UOS JOINT-project "A Cuban-Flemish Training and Research Program in Data Science and Big Data Analysis" (2018 - 2020) is gratefully acknowledged.

BIBLIOGRAPHY

Ayele, B. T., Lipkovich, I., Molenberghs, G., and Mallinckrodt, C. H. (2014). A multiple imputation based approach to sensitivity analysis and effectiveness assessments in longitudinal clinical trials. *Journal of Biopharmaceutical Statistics*, **24**, 211–228.

Box, G. E., and Tiao, G. C. (2011). *Bayesian inference in statistical analysis*. John Wiley & Sons.

Buuren, S. (2012). *Flexible imputation of missing data*. Chapman & Hall/CRC Press, London-New York.

Carpenter, J. R., and Kenward, M. G. (2013). *Multiple imputation and its application*. John Wiley & Sons, Ltd., Chichester, United Kingdom.

De Silva, A. P., Moreno-Betancur, M., De Livera, A. M., Lee, K. J., and Simpson, J. A. (2017). A comparison of multiple imputation methods for handling missing values in longitudinal data in the presence of a time-varying covariate with a non-linear association with time: a simulation study. *BMC Medical Research Methodology*, **17**, 1–11.

De Silva, A. P., Moreno-Betancur, M., De Livera, A. M., Lee, K. J., and Simpson, J. A. (2019). Multiple imputation methods for handling missing values in a longitudinal categorical variable with restrictions on transitions over time: a simulation study. *BMC Medical Research Methodology*, **19**, 1–14.

Díaz-Ordaz, K., Kenward, M. G., Cohen, A., Coleman, C. L., and Eldridge, S. (2014). Are missing data adequately handled in cluster randomised trials? A systematic review and guidelines. *Clinical Trials*, **II**, 590–600.

Diggle, P., and Kenward, M. G. (1994). Informative drop-out in longitudinal data analysis. *Appl. Stat.*, **43**, 49–93.

European Medicines Agency (EMA) (2010). *Guideline on Missing Data in Confirmatory Clinical Trials.* Available at http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2010/09/WC500096793.pdf.

Geman, S., and Geman, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **6**, 721–741.

Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, **57**, 97–109.

Kalaycioglu, O., Copas, A., King, M., and Omar, R. Z. (2016). A comparison of multiple-imputation methods for handling missing data in repeated measurements observational studies. *Journal of the Royal Statistical Society. Series A*, **179**, 683–706.

Little, R. J. A., and Rubin, D. B. (2014). *Statistical Analysis with Missing Data*. New York: John Wiley & Sons.

Liu, G. F., and Pang, L. (2015). On analysis of longitudinal clinical trials with missing data using reference-based imputation. *Journal of Biopharmaceutical Statistics*, DOI: 10.1080/10543406.2015.1094810.

Liu, G. F., and Pang, L. (2017). Control-based imputation and delta-adjustment stress test for missing data analysis in longitudinal clinical trials. *Statistics in Biopharmaceutical Research*, **9**, 186–194.

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

Molenberghs, G., and Kenward, M. G. (2007). *Missing Data in Clinical Studies*. Chichester: John Wiley & Sons.

Molenberghs, G., and Verbeke, G. (2005). *Models for Discrete Longitudinal Data*. New York: Springer.

National Academy of Sciences (NRC) (2010). *The Prevention and Treatment of Missing Data in Clinical Trials. Panel on Handling Missing Data in Clinical Trials.* Committee on National Statistics, Division of Behavioral and Social Sciences and Education, Washington, DC: The National Academies Press.

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

O'Kelly, M., and Ratitch, B. (2014). *Clinical Trials with Missing Data: A Guide for Practitioners*. New York: John Wiley & Sons.

Percy, D. F. (1992). Prediction for seemingly unrelated regressions. *Journal of the Royal Statistical Society. Series B (Methodological)*, **54**, 243–252.

Ratitch, B., O'Kelly, M., and Tosiello, R. (2013). Missing data in clinical trials: from clinical assumptions to statistical analysis using pattern mixture models. *Pharmaceutical Statistics*, **12**, 337–347.

Rubin, D. B. (1976). Inference and missing data. *Biometrika*, **63**, 581–592.

Rubin, D. B. (1978). Multiple imputations in sample surveys - a phenomenological Bayesian approach to nonresponse. *In Imputation and Editing of Faulty or Missing Survey Data*, 1–23. Washington, DC: U.S. Department of Commerce.

Rubin, D. B. (2004). *Multiple Imputation for Nonresponse in Surveys*. John Wiley & Sons.

SAS Institute Inc. (2011). **SAS OnlineDoc 9.3**. Cary, NC: SAS Institute Inc.

Schafer, J. L. (1997). Â€Analysis of incomplete multivariate data. Chapman & Hall/CRC.

Seaman, S. R., White, I. R., and Leacy, F. P. (2014). Comment on "Analysis of longitudinal trials with protocol deviations: a framework for relevant, accessible assumptions,

and inference via multiple imputation," by Carpenter, Roger, and Kenward. *Journal of Biopharmaceutical Statistics*, **24**, 1358–1362.

Tompsett, D. M., Leacy, F., Moreno-Betancur, M., White, I. R., and Heron, J. (2018). On the use of the not-at-random fully conditional specification (NARFCS) procedure in practice. *Statistics in Medicine*, 1–16.

Uranga, R., Allende, S., and Molenberghs, G. (2017). Simulation for the assessment of methods of treatment of longitudinal missing data in the context of clinical trials. *Operations Research*, **38**, 40–55.

Verbeke, G., and Molenberghs, G. (2000). *Linear Mixed Models for Longitudinal Data*. New York: Springer - Verlag.

Wood, A. M., White, I. R., and Thompson, S. G. (2004). Are missing outcome data adequately handled? A review of published randomized controlled trials in major medical journals. *Clinical trials*, **1**, 368–376.