

Multiple imputation by chained equations for systematically and sporadically missing multilevel data

Matthieu Resche-Rigon^{1,2,3} and Ian R White⁴

Statistical Methods in Medical Research
2018, Vol. 27(6) 1634–1649

© The Author(s) 2016

Reprints and permissions:

sagepub.co.uk/journalsPermissions.nav

DOI: 10.1177/0962280216666564

journals.sagepub.com/home/smm



Abstract

In multilevel settings such as individual participant data meta-analysis, a variable is ‘systematically missing’ if it is wholly missing in some clusters and ‘sporadically missing’ if it is partly missing in some clusters. Previously proposed methods to impute incomplete multilevel data handle either systematically or sporadically missing data, but frequently both patterns are observed. We describe a new multiple imputation by chained equations (MICE) algorithm for multilevel data with arbitrary patterns of systematically and sporadically missing variables. The algorithm is described for multilevel normal data but can easily be extended for other variable types. We first propose two methods for imputing a single incomplete variable: an extension of an existing method and a new two-stage method which conveniently allows for heteroscedastic data. We then discuss the difficulties of imputing missing values in several variables in multilevel data using MICE, and show that even the simplest joint multilevel model implies conditional models which involve cluster means and heteroscedasticity. However, a simulation study finds that the proposed methods can be successfully combined in a multilevel MICE procedure, even when cluster means are not included in the imputation models.

Keywords

Missing data, multilevel model, multiple imputation, chained equations, fully conditional specification, individual patient data meta-analysis

1 Introduction

Multiple imputation (MI) is a popular approach for handling the pervasive problem of missing data in biostatistics.¹ MI uses the distribution of the observed data to estimate a set of plausible values for the missing data, usually under a missing at random (MAR) assumption.² MI is a Bayesian procedure: given a joint prior distribution for the observed data and a specific data model, we obtain a posterior distribution of the missing values given the observed data. Missing data in a single variable are straightforward to impute, for example using an appropriate generalised linear model,^{3,4} but missing data in several variables present a more complex problem except on the rare occasions when data are monotone missing. There are two standard approaches. Firstly, a joint model may be assumed for the data. Most commonly, a multivariate normal model is assumed, as implemented in standard software.^{5–7} This is surprisingly effective despite being mis-specified for categorical data.⁸ Secondly, multiple imputation by chained equations (MICE), also known as multiple imputation by fully conditional specification, specifies a suitable conditional imputation model for each incomplete variable and iteratively imputes until convergence.^{3,9} The theoretical properties of MICE are not well understood: except in simple cases, conditional imputation models do not correspond to any joint model.^{10,11} Despite this it performs well

¹Service de Biostatistique et Information Médicale, Hôpital Saint-Louis, Paris, France

²Université Paris Diderot – Paris 7, Sorbonne Paris Cité, Paris, France

³ECSTRA Team, INSERM, Paris, France

⁴MRC Biostatistics Unit, Cambridge Institute of Public Health, Cambridge, UK

Corresponding author:

Matthieu Resche-Rigon, Service de Biostatistique et Information Médicale, Hôpital Saint-Louis, AP-HP, 1 avenue, Claude Vellefaux 75475, Paris Cedex 10, France.

Email: matthieu.resche-rigon@univ-paris-diderot.fr

in practice,^{12,13} especially when the conditional imputation models are well accommodated to the substantive model.¹⁴ MICE is also implemented in standard software.^{6,15,16}

Many datasets have a multilevel or clustered structure. Imputation that ignores this clustering leads to underestimation of the magnitude of clustering and hence underestimated standard errors, even if the analysis does allow for clustering.¹⁷ However, imputation that allows for the clustering through fixed effects of cluster overestimates the magnitude of the clustering and hence overestimates standard errors.^{18,19}

There is therefore a need for advanced imputation methods that allow for the clustering. Schafer and Yucel proposed a Gibbs sampler to generate MIs of continuous missing variables from a joint multivariate linear mixed model;^{20,21} their method is implemented in the PAN package.²² REALCOM software²³ and the R package *jomo*²⁴ extended this approach allowing missing data at any level and handling categorical data through latent normal variables. van Buuren proposed extending MICE to perform multilevel imputation by a Bayesian procedure.^{25,26} More recently, extensions were proposed to impute level 2 variables.²⁷

In this paper we focus on imputation of two-level clustered data, such as are found in individual participant data (IPD) meta-analysis²⁸ (where the cluster is the study) or in multi-centre studies (where the cluster is the centre). A feature of such data is that some variables may be ‘systematically missing’ – that is, missing for all individuals in one or more clusters.²⁹ This arises in particular in IPD meta-analysis when a potential confounding variable is not collected in one or more studies. Variables may also be ‘sporadically missing’ – that is, missing for some but not all individuals in one or more clusters.²⁹ Unfortunately, the multiple imputation methods and packages presented above are currently not able to handle systematically missing data except for *jomo*. Therefore, methods to handle systematically missing data have been proposed for continuous data²⁹ and recently extended to binary and discrete data.³⁰ These approaches handle the clustering by using generalized linear mixed models to impute data; they differ only in how they model the uncertainty around the between-cluster covariance parameters. However, they only deal with systematically missing data. The aim of this paper is to propose methods that simultaneously handle systematically and sporadically missing data.

The paper is organised as follows. Section 2 describes imputation of a single incomplete variable and makes two innovations. Firstly, it extends the method previously described for systematically missing data²⁹ to also handle sporadically missing data. Secondly, it proposes a new two-stage algorithm which conveniently allows for heteroscedasticity between clusters in the linear mixed model, as recommended by van Buuren in the sporadically missing case.²⁵ The two-stage algorithm also offers an easy extension to handle categorical data alongside normal data, although we here evaluate it only for normal data. In section 3, we explore the sorts of conditional models needed in multilevel MICE and explain why it is important to account for heteroscedasticity. In section 4, we evaluate the performance of the multilevel imputation algorithms in multilevel MICE. We illustrate the methods in an IPD meta-analysis in cardiovascular disease in section 5. Finally, section 6 provides a discussion.

2 Imputation of univariate missing data

Let y_i be a $n_i \times 1$ vector of observed outcomes on units $j \in \{1, \dots, n_i\}$ within cluster $i \in \{1, \dots, N\}$, $y_i = (y_{i1}, y_{i2}, \dots, y_{in_i})^T$. We consider the following linear mixed model

$$y_i = X_i\beta + Z_ib_i + e_i, \quad b_i \sim N(0, \Psi_b), \quad e_i \sim N(0, \Sigma_i) \quad (1)$$

where X_i is a $n_i \times p$ matrix of variables associated with the outcome y_i via β , a $p \times 1$ vector of fixed effects, and Z_i (typically equal to or contained within X_i) is a $n_i \times q$ matrix of variables associated with y_i via b_i , a $q \times 1$ vector of random effects. We model $b_i \sim N(0, \Psi_b)$ where Ψ_b is the $q \times q$ variance-covariance matrix of the random effects, and the $n_i \times 1$ error vector as $e_i \sim N(0, \Sigma_i)$ with $\Sigma_i = \sigma_i^2 I(n_i)$. Under this model, the posterior distribution of b_i given y_i is $b_i | y_i \sim N(m(y_i), v(y_i))$, with $m(y_i) = \Psi_b Z_i^T (Z_i \Psi_b Z_i^T + \Sigma_i)^{-1} (y_i - X_i \beta)$ and $v(y_i) = \Psi_b - \Psi_b Z_i^T (Z_i \Psi_b Z_i^T + \Sigma_i)^{-1} Z_i \Psi_b$.^{20,31,32}

Suppose $y = (y_1^T, y_2^T, \dots, y_N^T)^T$ contains missing data y^{mis} . y_i is systematically missing if it is missing for all individuals in cluster i ; it is sporadically missing if it is only partly incomplete.

Our goal is to generate independent draws under a MAR assumption from a posterior predictive distribution for the missing data $p(y^{mis} | y^{obs}) = \int p(y^{mis} | y^{obs}, \theta) p(\theta | y^{obs}) d\theta$ where $\theta = (\beta, \Psi_b, \{\sigma_i\})$ is the vector of parameters in

the linear mixed model (1) and $p(\theta|y^{obs})$ is the observed data posterior density of θ .¹ In practice, this may be achieved (with implicit vague priors) by:

- (1) fitting the model $p(y|\theta)$ to the units with observed y , yielding an estimate (typically an MLE) $\hat{\theta}$ with an estimated variance-covariance matrix S_{θ} ;
- (2) drawing a value of θ, θ^* , from its posterior, approximated as $N(\hat{\theta}, S_{\theta})$;
- (3) drawing values of y^{mis} from $p(y|\theta^*)$.

In the next two subsections, we propose two approaches to perform step 1. The first approach uses one-stage estimation of the model parameters.³³ The second approach, two-stage estimation, first estimates parameters within each cluster separately using the same model. Then, it applies multivariate meta-analysis methods using the summary statistics obtained at the first stage.³⁴ One-stage and two-stage approaches give similar results in individual patient data meta-analyses,³⁵ except when there are relatively few studies or the studies are small (especially with binary outcomes).³⁶ The two approaches both fit versions of model (1), but extra assumptions are convenient in each approach: in one-stage estimation we consider a homoscedastic model, i.e. a constant σ_i for all clusters, while in two-stage estimation we assume $Z_i = X_i$.

2.1 One-stage approach

This approach fits model (1) directly. To ease computation, we assume the residuals are homoscedastic between clusters: $\sigma_i^2 = \sigma^2 \forall i \in \{1, \dots, N\}$. Steps 1–3 are implemented as follows:

- (1) Estimate the model parameters, $\hat{\theta} = (\hat{\beta}, \hat{\Psi}_b, \hat{\sigma})$ and their variance covariance matrix S_{θ} by maximizing the restricted log-likelihood of model (1) fitted to the available data. We parameterise Ψ_b as proposed by Pinheiro and Bates,³⁷ using $\Psi_b^{trans} = ((\log \Psi_{brr} : r = 1 \text{ to } q), (\log \frac{1+\rho_{rs}}{1-\rho_{rs}} : r, s = 1 \text{ to } q, r < s))$ where $\rho_{rs} = \Psi_{brr}/\sqrt{\Psi_{brr}\Psi_{bss}}$. We assume that the estimates of the transformed parameters Ψ_b^{trans} follow a multivariate normal distribution.
- (2) Draw $\theta^* = (\beta^*, \Psi_b^{trans*}, \sigma^*)$ using $N(\hat{\theta}, \hat{S}_{\theta})$ and hence obtain Ψ_b^* and $\Sigma_i^* = \sigma^{*2}I(n_i)$. If Ψ_b^* is not positive semi-definite, make it positive semi-definite by setting negative eigenvalues to zero.³⁸
- (3) Compute $m^*(y_i)$ and $v^*(y_i)$, the posterior mean and variance of b_i , using Ψ_b^*, β^* and Σ_i^* :
 - (a) For clusters with systematically missing data, $m^*(y_i) = 0$ and $v^*(y_i) = \Psi_b^*$.
 - (b) For clusters with sporadically missing data, $m^*(y_i) = \Psi_b^{*T}Z_i(\Psi_b^{*T}Z_i^T + \Sigma_i^*)^{-1}(y_i - X_i\beta^*)$ and $v^*(y_i) = \Psi_b^* - \Psi_b^{*T}Z_i(\Psi_b^{*T}Z_i^T + \Sigma_i^*)^{-1}Z_i\Psi_b^*$.
 Then draw each b_i^* using $N(m^*(y_i), v^*(y_i))$ and for observation $j \in \{1, \dots, n_i\}$ with missing y_{ij} , draw $e_{ij}^* \sim N(0, \sigma^{*2})$ and set $y_{ij}^* = x_{ij}\beta^* + z_{ij}b_i^* + e_{ij}^*$.

This approach, recently extended by Jolani in a generalised linear model framework,³⁰ is close to van Buuren's method for sporadically missing data which used a Gibbs sampler to generate θ^* and b_i^* .²⁵ However, van Buuren argued that the imputation quality could be improved by allowing the within cluster variance to vary over the clusters, and modelled $\sigma_i \sim \frac{\sigma_0 \chi_{1/\phi}^2}{\phi}$, where σ_0 and ϕ are hyperparameters for the location of prior belief about residual variance and a measure of variability respectively.

Implementation

We fit the mixed model using the `lme()` function from the `nlme` package in R 3.1.1.^{39,40} This does not report the covariance between β^* and (ψ_b^*, σ^*) , so in practice we draw β^* and (ψ_b^*, σ^*) independently.

2.2 Two-stage approach

This approach uses procedures developed for multivariate meta-analysis.^{41,42} We first fit a separate model within each cluster and then combine the results using multivariate random-effects meta-analysis models. This approach naturally allows heteroscedastic level-1 variances. The model fitted in cluster i is

$$y_i = X_i\beta_i + e_i \quad (2)$$

with $e_i \sim N(0, \sigma_i^2 I(n_i))$. This is model (1) with $Z_i = X_i$ (i.e. every variable in the model has a random effect) and $\beta_i = \beta + b_i$; we propose how to allow $Z_i \neq X_i$ in the discussion. We approximate $\hat{\beta}_i \sim N(\beta_i, S_{\beta i})$ and $\log \hat{\sigma}_i \sim N(\log \sigma_i, S_{\sigma i})$. The between-studies model is $\beta_i \sim N(\beta, \Psi_\beta)$, where Ψ_β is the same as Ψ_b in section 2.1, and $\log \sigma_i \sim N(\log \sigma, \Psi_\sigma)$, where σ is an average within-cluster standard deviation. Then the marginal models are $\hat{\beta}_i \sim N(\beta, S_{\beta i} + \Psi_\beta)$ and $\log \hat{\sigma}_i \sim N(\log \sigma, S_{\sigma i} + \Psi_\sigma)$.

The main steps of the two-stage imputation procedure are:

- (1) (a) For each cluster i without systematically missing data, fit model (2) using y_i^{obs} (complete case analysis). This gives $\hat{\beta}_i$ with variance $S_{\beta i}$, and $\log \hat{\sigma}_i$ whose variance $S_{\sigma i}$ is obtained by noting that $\hat{\sigma}_i^{-2}$ follows a χ^2 distribution and using the delta method.
- (b) Perform a multivariate meta-analysis using the $\hat{\beta}_i$ and $S_{\beta i}$ from each cluster without systematically missing data. Using a REML estimator, obtain $\hat{\beta}$, $\hat{\Psi}_\beta$ and their variance-covariance matrices S_β and S_{Ψ_β} . Ψ_β is parameterised via its Cholesky decomposition Ψ_β^{chol} .
- (c) Perform a univariate meta-analysis using the $\hat{\sigma}_i$ and $S_{\sigma i}$ from each cluster without systematically missing data. Using a REML estimator, obtain $\log \hat{\sigma}$, $\hat{\Psi}_\sigma$ and their variance-covariance matrices S_σ and S_{Ψ_σ} .
- (2) Draw β^* using $N(\hat{\beta}, S_\beta)$ and Ψ_β^{chol*} using $N(\hat{\Psi}_\beta^{chol}, S_{\Psi_\beta^{chol}})$. Hence set $\Psi_\beta^* = \Psi_\beta^{chol*}(\Psi_\beta^{chol*})^T$: the use of the Cholesky decomposition ensures a positive semi-definite Ψ_β^* . Similarly draw $\log \sigma^*$ and Ψ_σ^* using $N(\log \hat{\sigma}, S_\sigma)$ and $N(\hat{\Psi}_\sigma, S_{\Psi_\sigma})$.
- (3) Draw the missing y_{ij} :
 - (a) For clusters with systematically missing data, draw β_i^* using $N(\beta^*, \Psi_\beta^*)$ and draw σ_i^* using $\log \sigma_i^* \sim N(\log \sigma^*, \Psi_\sigma^*)$.
 - (b) For clusters with sporadically missing data, draw β_i^* from its posterior given Ψ_β^* and β^* ,

$$\beta_i^* \sim N\left((\Psi_\beta^{*-1} + S_{\beta i}^{-1})^{-1}(\Psi_\beta^{*-1}\beta^* + S_{\beta i}^{-1}\hat{\beta}_i), (\Psi_\beta^{*-1} + S_{\beta i}^{-1})^{-1}\right)$$

and draw $\log \sigma_i^*$ from its posterior given Ψ_σ^* and $\log \sigma^*$

$$\log \sigma_i^* \sim N\left(\frac{\log \sigma^* / \Psi_\sigma^* + \log \hat{\sigma}_i / S_{\sigma i}}{1 / \Psi_\sigma^* + 1 / S_{\sigma i}}, \frac{1}{1 / \Psi_\sigma^* + 1 / S_{\sigma i}}\right).$$

- (c) For each missing y_{ij} , draw $e_{ij}^* \sim N(0, \sigma_i^{*2})$ and set $y_{ij}^* = x_{ij}\beta_i^* + e_{ij}^*$.

The REML estimation proposed in step 1 can be computationally slow. We therefore propose the method of moments (MM) as a simple and computationally efficient alternative. In step 1(c), we use the univariate MM described in Dersimonian and Laird.⁴³ In step 1(b), we use the multivariate extension proposed by Jackson et al.³⁸ The MM does not allow estimation of S_{Ψ_β} and S_{Ψ_σ} , so we are forced to modify Step 2 by setting $\Psi_\beta^* = \hat{\Psi}_\beta$ and $\sigma^* = \hat{\sigma}$.

Implementation

Multivariate meta-analysis is performed using the *mvmeta* package in R 3.1.1.^{34,44,45}

3 Imputing two or more incomplete variables

3.1 MICE

It is usual for missing values to occur in several variables. MICE specifies the multivariate imputation model on a variable-by-variable basis by a set of conditional densities obtained by different regression models, one for each incomplete variable.⁴⁶ Thus, MICE allows the imputation model to be adapted to the type of each variable. Conditional distributions for the missing data given the other data for each incomplete variable are iteratively obtained, missing values being replaced by simulated draws. The algorithm may be initiated by a simple random sampling with replacement from the observed values. Once each variable has been imputed the complete cycle is repeated at least 5–10 times to stabilize the posterior distribution obtained for each variable.³ In practice, posterior distributions could be derived either from maximum likelihood estimators or using Gibbs samplers as proposed by van Buuren.²⁵ In the multilevel setting, the use of MICE was first proposed by van Buuren²⁶ and was applied for systematically missing data by Resche-Rigon et al.²⁹ We propose using the methods described in section 2 to provide the conditional imputations needed at each step of the MICE procedure.

3.2 Conditional models in the multilevel setting

Ideally, one wants to use conditional imputation models compatible with the unknown overall data model. For complex models this is often unachievable. We use a simple joint model to explore mathematically the difficulties arising in specifying conditional models in the multilevel setting. We consider two incomplete variables x_{1ij} and x_{2ij} , with the joint model

$$\begin{pmatrix} x_{1ij} \\ x_{2ij} \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} + \begin{pmatrix} u_{1i} \\ u_{2i} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1ij} \\ \varepsilon_{2ij} \end{pmatrix} \quad (3)$$

where $\begin{pmatrix} u_{1i} \\ u_{2i} \end{pmatrix} \stackrel{\text{iid}}{\sim} N(0, \Psi_u)$ and $\begin{pmatrix} \varepsilon_{1ij} \\ \varepsilon_{2ij} \end{pmatrix} \stackrel{\text{iid}}{\sim} N(0, \Sigma_\varepsilon)$.

We prove in Appendix 1 that the conditional expectation of x_{2ij} depends on x_{1ij} and \bar{x}_{1i} , the mean of x_{1ij} within cluster i , and on n_i , the size of cluster i (equation (8)). Similarly, the conditional variance of x_{2ij} depends on n_i (equation (9)). Appendix 2 generalises this result to the case of many incomplete variables by letting x_{1ij} and x_{2ij} be vectors. These results suggest (1) that we should incorporate the cluster means of the predictors in the imputation model and (2) that if cluster sizes vary then the level-1 variance should vary between clusters, i.e. that the imputation model should be heteroscedastic. Thus, a simple joint model implies complex conditional models: this is unlike the situation with independent observations, when a joint normal model implies simple conditional models.¹¹

Many settings, including the simulation setting of section 4, include random slopes: for example, we might combine model (3) with a random slopes model regressing y_{ij} on x_{1ij} and x_{2ij} . Under this joint model, the cluster-specific density of (x, y) is multivariate normal, but the marginal density is not. Deriving the conditional distributions in this model is intractable, and we anticipate that the problems identified above will be joined by others.

4 Simulation study

The methods proposed in section 2 make certain approximations, and section 3 suggests that combining them into a MICE procedure involves mis-specification of the conditional models. We therefore conduct a simulation study aiming (a) to explore and compare the performance of the methods in a MICE procedure, and (b) to explore whether their performance is improved by including cluster means or heteroscedasticity in the imputation model.

4.1 Simulating complete data

As in section 2, let y_i, x_{1i}, x_{2i} denote $n_i \times 1$ vectors containing outcomes and covariates respectively on units $j \in \{1, \dots, n_i\}$ within cluster $i \in \{1, \dots, N\}$. The covariates were drawn from a possibly heteroscedastic version of model (3): $(x_{1ij}, x_{2ij})^T \sim N(\mu_i, \Sigma_{\varepsilon i})$ and $\mu_i \sim N(\mu, \Psi_u)$. We allowed the marginal variances to vary across clusters using $\log \Sigma_{\varepsilon i k k} \sim N(\log \Sigma_{\varepsilon, k}, \Psi_{\Sigma_{\varepsilon, k}})$ for $k = 1, 2$, but we fixed the marginal correlations $\Sigma_{\varepsilon i 12} / \sqrt{\Sigma_{\varepsilon i 11} \Sigma_{\varepsilon i 22}} = \rho_\varepsilon$. Our parameter values for the base case were $\mu = (0, 0)^T$, $\Psi_u = 0.5^2 \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}$, $\Sigma_{\varepsilon, 1} = \Sigma_{\varepsilon, 2} = 0.5^2$, $\rho_\varepsilon = 0.5$ and $\Psi_{\Sigma_{\varepsilon, 1}} = \Psi_{\Sigma_{\varepsilon, 2}} = 0$, giving homoscedasticity with $\Sigma_{\varepsilon i} = \Psi_u$ for all i and hence intra-cluster correlations of approximately 0.5 for both x_1 and x_2 .

The outcome data y_{ij} were generated from a random intercepts and slopes model

$$y_{ij} = \beta_0 + \beta_1 x_{1ij} + \beta_2 x_{2ij} + b_{0i} + b_{1i} x_{1ij} + b_{2i} x_{2ij} + e_{ij} \quad (4)$$

with $e_{ij} \sim N(0, \sigma_i^2)$, $\log \sigma_i \sim N(\log \sigma, \Psi_\sigma)$ and $b_i = (b_{0i}, b_{1i}, b_{2i}) \sim N(0, \Psi_b)$. Our parameter values for the base case were $\beta_0 = 0$, $\beta_1 = 0.5$ and $\beta_2 = 1$; Ψ_b had all diagonal elements 0.5^2 and all correlations 0.3; $\sigma = 0.5$ and $\Psi_\sigma = 0$, so the outcome model was homoscedastic with intra-cluster correlation (conditional on $x_{1ij} = x_{2ij} = 0$) of 0.5.

Following Appendices 1 and 2, $E(x_{1ij} | \mathbf{x}_{2ij}, y_{ij})$ and $E(x_{2ij} | \mathbf{x}_{1ij}, y_{ij})$ may depend on \bar{y}_i and on \bar{x}_{2i} and \bar{x}_{1i} respectively. We estimate this dependency using a large simulated data set of 1000 clusters of size 100.

Estimated coefficients of \bar{y}_i and \bar{x}_{2i} in the model for x_{1ij} and of \bar{y}_i and \bar{x}_{1i} in the model for x_{2ij} were -0.020 , 0.049 , 0.007 and 0.020 respectively, much smaller than the coefficients of y_i , x_{1i} , x_{2i} which were all greater than 0.12 .

The total number of patients was fixed at 2000. The number of clusters was $N = 20$ yielding $n_i = 100$ patients per cluster. A total of 1000 independent datasets were generated for each setting.

4.2 Simulating missing data

y_i was complete for all i . x_{1i} and x_{2i} were independently systematically missing with probability π_{sys} . For any cluster where x_{1i} or x_{2i} was not systematically missing, x_{1ij} and x_{2ij} were independently sporadically missing with probability π_{spor} . This is a missing completely at random mechanism. Our parameter values were $\pi_{sys} = 0.2$ and $\pi_{spor} = 0.3$ in all cases, leading to an overall probability of missing data equal to 0.44 .

4.3 Imputing missing data

Under this data generating model, the conditional distributions $[\mathbf{x}_{1i}|\mathbf{x}_{2i}, \mathbf{y}_i]$ where $\mathbf{x}_{1i} = (x_{1i1}^T, x_{1i2}^T, \dots, x_{1in_i}^T)^T$ etc. are not of the simple forms given in equations (1) and (2). Nevertheless, in order to explore the practical consequences of imputation model misspecification, missing data were imputed using a MICE algorithm, using as imputation model either homoscedastic equation (1) (for the one-stage approach) or heteroscedastic equation (2) (for the two-stage approach). The two-stage approach was implemented using either REML or the MM in steps 1(b) and 1(c). When imputing \mathbf{x}_1 , the outcome in equation (1) or (2) was \mathbf{x}_1 , and the covariates were \mathbf{x}_2 and \mathbf{y} , with both fixed and random effects. A similar model was used to impute \mathbf{x}_2 . To evaluate the impact of the cluster means within the imputation model we used two series of imputation models, either excluding or including the cluster means. Cluster means were included in the one-stage method by adding them as predictors in the imputation models, and in the two-stage method by using meta-regression in the second stage with cluster means as predictors. The incomplete data were imputed $m = 5$ times using 10 iterations of the chained equations algorithm.

4.4 Analysis

In every case, the analysis model was equation (4). The data were analysed before data deletion as a benchmark for the multiple imputation procedure, and after data deletion using complete cases. After imputation, the model was estimated in each imputed data set using the REML estimator of the `lmer()` function of the *lme4* R-package⁴⁷ and the estimates were combined using Rubin's rules.¹ The performance of each method was assessed by computing the empirical mean of the parameter estimates, root mean square estimated standard error (Model SE), empirical Monte Carlo standard error (Emp SE), the coverage of nominal 95% confidence intervals (95%CI), and the mean run time per simulated data set.

4.5 Results

In the base case (Table 1), point estimates are slightly negatively biased for $\beta_1 = 0.5$ and unbiased for $\beta_2 = 1$ for two-stage methods. Empirical standard errors are lower for multiple imputation approaches than for complete case analysis, reflecting the better use made of the data. Model-based standard errors are below empirical standard errors, in particular for the one-stage approach. The underestimated standard errors appear to arise from underestimated random effect variances. Moderate under-coverage is observed, and is only partly explained by the underestimated standard errors: only two-stage methods achieve coverage within 5% of the nominal 95%. Mean run times are in favour of the two-stage MM especially compared to the two-stage REML. Including cluster means in imputation models has little impact on results (lower part of Table 1).

We next modified the data generating mechanism to explore performance when the coefficients of cluster means are clearly different from 0. To do this we set $\Sigma_{\epsilon,1} = 0.5^2$, $\Sigma_{\epsilon,2} = 0.17^2$ and $\Psi_u = \begin{pmatrix} 0.17^2 & 0.5^2 \times 0.17 \\ 0.5^2 \times 0.17 & 0.5^2 \end{pmatrix}$. Using a large simulated data set, the estimated coefficients of \bar{y}_i and \bar{x}_{2i} in the model for x_{1ij} and of \bar{y}_i and \bar{x}_{1i} in the model for x_{2ij} were -0.16 , -0.87 , 0.22 and 0.95 respectively, larger than the corresponding coefficients of y_i , x_{1i} , x_{2i} which were 0.30 , 0.60 , 0.07 and 0.12 , respectively.

Table 1. Base-case simulations. $\Psi_u = 0.5^2 \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}$, $\Sigma_{\varepsilon,1} = \Sigma_{\varepsilon,2} = 0.5$, $\Psi_{\Sigma_{\varepsilon,1}} = \Psi_{\Sigma_{\varepsilon,2}} = 0$.

Method	β_1				$\sqrt{\Psi_{b11}}$	β_2				$\sqrt{\Psi_{b22}}$	Mean run time
	Est mean	Model SE	Emp SE	Cover	Est mean	Est mean	Model SE	Emp SE	Cover	Est mean	
True value	0.5				0.5	1				0.5	
No missing data	0.501	0.113	0.115	0.926	0.489	1.002	0.114	0.117	0.922	0.494	
Complete case	0.505	0.145	0.154	0.925	0.483	1.005	0.146	0.150	0.916	0.486	
Imputation methods, excluding cluster mean from imputation model											
One-stage	0.485	0.112	0.129	0.899	0.407	0.984	0.105	0.132	0.874	0.410	1 min 51 s
Two-stage	0.488	0.125	0.134	0.919	0.461	1.008	0.121	0.134	0.915	0.474	3 min 57 s
Two-stage MM	0.489	0.124	0.136	0.908	0.456	1.010	0.121	0.131	0.910	0.472	1 min 3 s
Imputation methods, including cluster mean in imputation model											
One-stage	0.486	0.112	0.129	0.888	0.406	0.982	0.105	0.134	0.868	0.409	3 min 26 s
Two-stage	0.485	0.125	0.137	0.895	0.469	1.005	0.122	0.135	0.909	0.476	4 min 11 s
Two-stage MM	0.487	0.122	0.137	0.906	0.464	1.006	0.120	0.133	0.915	0.474	1 min 6 s

Est mean: mean estimate over imputed data sets; Model SE: standard error derived from Rubin's rules (mean over imputed data sets); Emp SE: standard deviation of estimate over imputed data sets; Cover: coverage of nominal 95% confidence interval.

Table 2. Simulations when the coefficients of cluster means are clearly different from 0. $\Psi_u = \begin{pmatrix} 0.17^2 & 0.5^2 \times 0.17 \\ 0.5^2 \times 0.17 & 0.5^2 \end{pmatrix}$, $\Sigma_{\varepsilon,1} = 0.5^2$, $\Sigma_{\varepsilon,2} = 0.17^2$, $\Psi_{\Sigma_{\varepsilon,1}} = \Psi_{\Sigma_{\varepsilon,2}} = 0$.

Method	β_1				$\sqrt{\Psi_{b11}}$	β_2				$\sqrt{\Psi_{b22}}$	Mean run time
	Est mean	Model SE	Emp SE	Cover	Est mean	Est mean	Model SE	Emp SE	Cover	Est mean	
True value	0.5				0.5	1				0.5	
No missing data	0.500	0.113	0.112	0.934	0.494	1.003	0.131	0.132	0.929	0.485	
Complete case	0.500	0.146	0.148	0.926	0.486	1.002	0.188	0.191	0.934	0.484	
Imputation methods, excluding cluster mean from imputation model											
One-stage	0.475	0.110	0.125	0.897	0.416	1.036	0.183	0.174	0.928	0.521	2 min 54 s
Two-stage	0.486	0.120	0.129	0.919	0.465	1.014	0.204	0.174	0.960	0.631	3 min 52 s
Two-stage MM	0.491	0.121	0.130	0.927	0.463	1.011	0.215	0.176	0.963	0.667	1 min 3 s
Imputation methods, including cluster mean in imputation model											
One-stage	0.474	0.108	0.125	0.893	0.412	1.025	0.179	0.172	0.942	0.514	5 min 19 s
Two-stage	0.481	0.119	0.129	0.915	0.465	1.016	0.196	0.177	0.951	0.627	4 min 12 s
Two-stage MM	0.489	0.118	0.129	0.916	0.465	0.995	0.204	0.181	0.959	0.655	1 min 5 s

Note: Abbreviations as in Table 1.

The bias in β_1 and β_2 remains small, but seems smaller for two-stage approaches (Table 2). The MI methods underestimate Ψ_{b11} (especially with one-stage approaches) and overestimate Ψ_{b22} (especially with two-stage approaches), leading to mis-estimated model-based standard errors. In this setting, in which cluster means could have an impact on the efficiency of the imputation model, including cluster means has very little effect on the final results, with only a small reduction in model-based standard errors.

We next modified the data generating mechanism of Table 2 by introducing unequal cluster sizes, with 10 clusters of size 160 and 10 clusters of size 40. Results (Table 3) were very close to those obtained in Table 2, and two-stage methods seem to be less biased than one-stage approach.

Finally, Table 4 reports a simulation study using the same parameter values as in Table 2, but with heteroscedasticity on X_1 and X_2 introduced by setting $\Psi_{\Sigma_{\varepsilon,1}} = 0.17^2$, $\Psi_{\Sigma_{\varepsilon,2}} = 0.06^2$. Results were still less biased

Table 3. Simulations when the coefficients of cluster means are clearly different from 0 and cluster sizes are unequal (10 clusters of size 160, 10 clusters of size 40). $\Psi_u = \begin{pmatrix} 0.17^2 & 0.5^2 \times 0.17 \\ 0.5^2 \times 0.17 & 0.5^2 \end{pmatrix}$, $\Sigma_{\varepsilon,1} = 0.5^2$, $\Sigma_{\varepsilon,2} = 0.17^2$, $\Psi_{\Sigma_{\varepsilon,1}} = \Psi_{\Sigma_{\varepsilon,2}} = 0$.

Method	β_1				$\sqrt{\Psi_{b11}}$	β_2				$\sqrt{\Psi_{b22}}$	Mean run time
	Est mean	Model SE	Emp SE	Cover	Est mean	Est mean	Model SE	Emp SE	Cover	Est mean	
True value	0.5				0.5	1				0.5	
No missing data	0.495	0.115	0.115	0.946	0.492	0.997	0.135	0.139	0.950	0.485	
Complete case	0.494	0.147	0.155	0.935	0.483	1.002	0.192	0.206	0.912	0.495	
Imputation methods, excluding cluster mean from imputation model											
One-stage	0.473	0.111	0.132	0.900	0.415	1.023	0.185	0.183	0.941	0.513	3 min 13 s
Two-stage	0.481	0.124	0.135	0.925	0.468	0.994	0.213	0.185	0.971	0.632	3 min 53 s
Two-stage MM	0.486	0.123	0.136	0.925	0.466	0.996	0.221	0.193	0.968	0.668	1 min 3 s
Imputation methods, including cluster mean in imputation model											
One-stage	0.473	0.110	0.131	0.899	0.412	1.009	0.181	0.183	0.947	0.504	5 min 42 s
Two-stage	0.479	0.123	0.134	0.928	0.472	0.998	0.206	0.187	0.968	0.637	4 min 15 s
Two-stage MM	0.482	0.121	0.135	0.920	0.469	0.982	0.213	0.186	0.974	0.664	1 min 5 s

Note: Abbreviations as in Table 1.

with two-stage methods than with the one-stage approach. Ψ_{b22} was largely overestimated. Model-based standard errors obtained with the two-stage approaches were overestimated.

5 Application to GREAT data

The GREAT Network explored risk factors associated with short-term mortality in acute heart failure (AHF). Their dataset consisted of 12 observational cohorts: 6 were based in Western Europe (2 in Italy and 1 in each of France, Finland, Switzerland, and Spain), 2 in Central Europe (Austria, Czech Republic), 2 in the United States, 1 in Asia (Japan), and 1 in Africa (Tunisia).⁴⁸ The principal investigators of each cohort study submitted the original data collected for each patient, including a list of patient characteristics and potential risk factors.⁴⁹

The biomarker of interest was brain natriuretic peptide (BNP), which is known to be elevated in acute heart failure. Because our aim is to develop and compare statistical methods, we take as outcome the left ventricular ejection fraction (LVEF) measured by echocardiography, a potential sign of AHF. Our aim is to construct a model predicting the value of LVEF using BNP and six other variables: gender, body mass index (BMI), age, systolic blood pressure (SBP), diastolic blood pressure (DBP) and heart rate (HR). The analysis model is a linear mixed model with all predictors included as linear terms. The intercept and the slope for BNP were both random at cohort level. A description of the variables considered is given in Table 5. BNP was log-transformed for analysis.

This analysis was complicated by missing data (Table 6). BNP measurement is a relatively recent innovation, so BNP was systematically missing in 4 cohorts and sporadically missing in 5 others. BMI, SBP and DBP were also systematically missing in some cohorts and sporadically missing in others, while HR and age were only sporadically missing. Gender and LVEF were complete by design.

The results (Table 7) show that multiple imputation approaches have a substantial impact on point estimates of the coefficients. We observe a gain in precision for all MI methods for complete variables (gender, age) and for most incomplete variables (BMI, SBP, DBP, HR). Only log(BNP) has a bigger standard error after imputation, perhaps reflecting a difference between the smaller cohorts with systematically missing data and the more complete cohorts. Results from different imputation methods agree except for log(BNP), whose coefficient and standard errors vary substantially.

6 Discussion

There is increasing need for ways to handle missing values in multilevel structures, notably with the development of meta-analysis of IPD.⁵⁰ IPD, whether from randomised clinical trials or observational studies, have the

Table 4. Simulations when the coefficients of cluster means are clearly different from 0 and with heteroscedasticity on X_1 and X_2 .

$$\Psi_u = \begin{pmatrix} 0.17^2 & 0.5^2 \times 0.17 \\ 0.5^2 \times 0.17 & 0.5^2 \end{pmatrix}, \Sigma_{\varepsilon,1} = 0.5^2, \Sigma_{\varepsilon,2} = 0.17^2, \Psi_{\Sigma_{\varepsilon,1}} = 0.17^2, \Psi_{\Sigma_{\varepsilon,2}} = 0.06^2.$$

Method	β_1				$\sqrt{\Psi_{b11}}$	β_2				$\sqrt{\Psi_{b22}}$	Mean run time
	Est mean	Model SE	Emp SE	Cover	Est mean	Est mean	Model SE	Emp SE	Cover	Est mean	
True value	0.5				0.5	1				0.5	
No missing data	0.500	0.113	0.115	0.943	0.490	1.000	0.134	0.136	0.946	0.488	
Complete case	0.503	0.146	0.147	0.943	0.483	0.988	0.187	0.203	0.918	0.484	
Imputation methods, excluding cluster mean from imputation model											
One-stage	0.451	0.115	0.122	0.912	0.425	1.002	0.204	0.185	0.961	0.598	1 min 41 s
Two-stage	0.484	0.139	0.131	0.951	0.505	1.019	0.268	0.200	0.993	0.796	3 min 54 s
Two-stage MM	0.500	0.136	0.131	0.956	0.499	1.018	0.260	0.200	0.988	0.796	1 min 4 s
Imputation methods, including cluster mean in imputation model											
One-stage	0.453	0.115	0.120	0.922	0.422	0.995	0.200	0.182	0.963	0.582	3 min 46 s
Two-stage	0.482	0.138	0.129	0.953	0.509	1.017	0.262	0.201	0.987	0.802	4 min 10 s
Two-stage MM	0.497	0.133	0.128	0.941	0.502	1.001	0.252	0.195	0.986	0.787	1 min 6 s

Note: Abbreviations as in Table 1.

Table 5. GREAT study: description of variables.

Variables	Values	Patients (N = 4546)	Statistics
Gender	0	1767	38.87%
	1	2779	61.13%
Age		4540	73.28 [63.75;80]
BMI		3298	27.37 [24.34;31.06]
SBP		4185	136 [116;160]
DBP		4170	80 [69;90]
HR		4439	88 [73;105]
BNP		2034	973 [471.2;1860]
LVEF		4546	40 [27;55]

Note: Categorical variables are presented by counts and percentages and quantitative variable by their median and quartiles.

BMI: body mass index; SBP: systolic blood pressure; DBP: diastolic blood pressure; HR: heart rate; BNP: brain natriuretic peptide; LVEF: left ventricular ejection fraction.

Table 6. GREAT study: percentages of missing values by variables and cohort.

Cohort	1	2	3	4	6	7	8	9	10	11	12	13
Sample size	410	567	210	375	107	267	203	354	137	48	1790	78
Percentage missing												
Gender	0	0	0	0	0	0	0	0	0	0	0	0
Age	0	0	0	0	0	0	0	1	0	0	0	0
BMI	36	19	43	2	1	100	0	44	1	100	21	60
SBP	1	2	1	1	0	100	1	16	0	0	1	0
DBP	2	3	2	1	0	100	2	16	0	0	1	0
HR	3	1	1	2	0	0	1	19	0	4	0	0
BNP	57	1	0	4	100	100	0	12	0	100	93	100
LVEF	0	0	0	0	0	0	0	0	0	0	0	0

Note: Abbreviations as in Table 5.

Table 7. GREAT study: results from linear mixed model analysis.

	Gender		Age		BMI		SBP		DBP		HR		log(BNP)	
	β	SE	β	SE	β	SE	β	SE	β	SE	β	SE	β	SE
Complete case	−6.869	0.803	0.200	0.030	0.184	0.064	0.167	0.016	−0.176	0.028	−0.060	0.016	−9.887	1.176
Imputation methods, excluding cohort mean in imputation model														
One-stage	−5.669	0.475	0.165	0.019	0.233	0.043	0.124	0.012	−0.124	0.023	−0.048	0.009	−10.359	1.112
Two-stage	−5.795	0.474	0.154	0.019	0.226	0.048	0.129	0.012	−0.132	0.023	−0.050	0.009	−9.313	2.034
Two-stage MM	−5.824	0.463	0.156	0.020	0.243	0.043	0.127	0.010	−0.127	0.017	−0.051	0.009	−9.816	1.900
Imputation methods, including cohort mean in imputation model														
One-stage	−5.662	0.503	0.170	0.020	0.230	0.049	0.123	0.011	−0.123	0.019	−0.049	0.009	−10.096	1.666
Two-stage	−5.737	0.455	0.158	0.020	0.243	0.045	0.125	0.010	−0.125	0.020	−0.050	0.009	−6.440	2.803
Two-stage MM	−5.950	0.484	0.152	0.020	0.250	0.047	0.137	0.013	−0.141	0.023	−0.049	0.009	−6.302	2.672

Note: Abbreviations as in Table 5

advantage of facilitating consistent definitions of outcomes, exposures and confounders and consistent analyses between studies. Nevertheless, the availability of confounders typically varies between studies. In this paper, we proposed a method to handle both systematically and sporadically missing covariates in a two-level structure using MICE. We explored the conditional imputation models needed in multilevel MICE and showed that cluster means and heteroscedasticity should be considered in the imputation model. We proposed a new two-stage approach in which we first fit the imputation model within each cluster and then combine the results using multivariate random effects models. Our simulation studies showed broadly good performance for the MICE methods, a very low impact of including the cluster means, and an advantage for heteroscedastic models. Small biases and slight under-coverage were observed in the presence of only systematically missing data.²⁹ These biases disappeared when only sporadically missing covariates were considered, suggesting that they are linked to the presence of systematically missing covariates (Appendix 3). Moreover, the two-stage approach using the MM was more computationally efficient.

Our work extended a previous algorithm that we proposed to handle systematically missing data.²⁹ In recent work, Jolani et al. also extended our one-stage approach to handle systematically missing binary data.³⁰ The approaches differ in the way uncertainty is introduced around the estimated between-cluster covariance parameters, whose sampling distribution is skewed: we used a general log-transformation for the variances and Fisher's transformation for correlations, as proposed by Pinheiro and Bates³⁷ and applied in some current software packages for fitting non-linear mixed models (e.g. nlme in R), whereas Jolani et al. used a Bayesian framework with a Wishart distribution for the inverse between-cluster covariance.³⁰

To our knowledge, no other package can handle both systematically and sporadically missing data except the recent *jomo* R-package.^{24,51} Other methods developed to impute one type of missing data could be modified to handle the other type. In particular, the *2l.norm* routine¹⁵ for sporadically missing data in MICE R-package could be adapted using step 3(a) and 3(b) of the one stage approach, although using a MCMC algorithm at each step of the chained equations algorithm seems to be computationally laborious.³⁰

Another alternative is Gibbs sampling. Schafer proposed a Gibbs sampler to generate multiple imputations of a single continuous incomplete variable from a joint multivariate linear mixed model;²⁰ this approach was implemented in the PAN package²² and also in the MICE package.¹⁵ The REALCOM package²³ and the recent *jomo* R-package²⁴ perform multilevel joint modelling multiple imputations, handling binary and categorical variables through latent normal variables. The *jomo* R-package is also able to use heteroscedastic framework, and should be in that sense close to our two-stage approach.

The multilevel MICE approach has several problems, especially the heteroscedasticity of the conditional imputation models demonstrated in section 3.2. Our simulation studies showed that with heteroscedastic models our two-stage methods (which naturally allow heteroscedasticity) outperform the one-stage method. It would be possible to introduce heteroscedasticity in the one-stage model, thus avoiding any issues due to small studies or small number of studies,³⁶ but in our current view this is computationally laborious and unrealistic within a MICE procedure. To our knowledge only the *2l.norm* routine in the MICE package allows for heteroscedasticity.²⁵ More recently, Yucel described a heteroscedastic imputation model for imputing

multivariate multilevel continuous data.²¹ Similarly, Jolani argued for a cluster-specific error variance which they related to cluster-level characteristics.³⁰

Another potential problem is compatibility between the full conditional distribution of each incomplete variable (represented by the specification of variable-by-variable imputation models) and the global joint distribution of the multivariate missing data. Ideally, the conditional distribution should be derived from a joint probability distribution.^{26,52} Recent studies have identified compatibility in simple cases, notably for the linear model without any random effect,¹⁰ but the MICE approach is usually validated only by simulation studies, and the results appear to be robust even with incompatible models.^{26,53} In this work, we studied simulations with only two missing covariates but extension of our results with more missing variables should be valid provided that conditional imputation models are well specified. Moreover, having a valid joint distribution may be “less important than incorporating information from other variables and unique features of the dataset”.⁵⁴ This could partially explain the good results observed for our method even without including the cluster means in the imputation model. Another possible explanation is that \bar{x}_{1i} and \bar{x}_{2i} are normally distributed and can therefore be absorbed into the random intercept term.

The number of random effects that we can consider is a further problem, especially when most studies have systematically missing data. We previously showed that random effects on the intercept and on the factor of interest could be sufficient to produce good results.²⁹ A large number of random effects can make the computational time prohibitive. Nevertheless, we prefer to put a random effect on all the variables, and the two-stage method with the simple and computationally efficient MM is a good alternative despite the fact that we are forced to ignore uncertainty in the variance components Ψ_{β}^* and σ^* . The computational advantage becomes much greater as the number of random effects increases. Of course, our methods could benefit from parsimonious modelling of the variance-covariance matrix of random effects in the imputation model (e.g. modelling independent random effects). One could also constrain some components of the random effect covariance matrix to zero, based on empirical covariance matrix estimates.²³ Such constraints could be directly implemented in the two-stage methods by forcing some between-studies variances to be zero in the second step. In practice, this corresponds to considering a smaller set of variables in Z_i than in X_i . The two stage approach also allows level-2 variables to be included through meta-regression in the second stage.

An important extension of our imputation model is to impute categorical variables. In the past, this has been done in a MICE framework using random-effects logistic imputation models,³⁰ and in a joint model framework using latent continuous normally distributed variables.²⁴ Our two-stage approach would be particularly simple to modify for any generalised linear or other regression model, and this is the subject of future work.

In conclusion, MICE seems to be a valuable approach to handle both systematically and sporadically missing data. We proposed three methods, but only the two-stage methods easily take account of the potential heteroscedasticity in the imputation model, and they outperform the one-stage approach in some settings. The two-stage methods perform quite similarly, but the computational time needed to fit a REML estimator is a disadvantage. Therefore, the two-stage method with the MM approach appears to be a solid and efficient alternative.

Acknowledgements

We thank the Global Research on Acute Conditions Team (GREAT, <http://www.greatnetwork.org>) for allowing the use of their data.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: MR-R was supported by Agence Nationale de Sécurité du Médicament et des Produits de Santé. IRW was supported by the Medical Research Council [Unit Programme number U105260558].

References

1. Rubin DB. *Multiple Imputation for Nonresponse in Surveys*, Wiley Series in Probability and Statistics. New York: Wiley, 1987.
2. Little R and Rubin D. *Statistical analysis with missing data*, 2nd ed. Hoboken, NJ: Wiley, 2002.
3. White I, Royston P and Wood A. Multiple imputation using chained equations: issues and guidance for practice. *Stat Med* 2011; **30**: 377–399.
4. Carpenter J and Kenward M. *Multiple imputation and its application*. New York: John Wiley & Sons, 2012.
5. Novo A and Schafer J. Package NORM, R package version 1.0-9.5, 2015, <http://CRAN.R-project.org/package=norm>
6. StataCorp. *Stata multiple-imputation reference manual: release 12*. College Station, TX: Stata Press, 2011.
7. SAS Institute Inc. *SAS/STAT 9.1 user's guide*. Chapter 46. Cary, NC: SAS Institute Inc., 2004.
8. Schafer J. *Analysis of incomplete multivariate data*. Boca Raton: Chapman & Hall/CRC, 1997.
9. van Buuren S, Boshuizen H and Knook D. Multiple imputation of missing blood pressure covariates in survival analysis. *Stat Med* 1999; **18**: 681–694.
10. Chen H. Compatibility of conditionally specified models. *Stat Probab Lett* 2010; **80**: 670–677.
11. Hughes R, White I, Seaman S, et al. Joint modelling rationale for chained equations. *BMC Med Res Meth* 2014; **14**: 28.
12. van Buuren S, Brand J, Groothuis-Oudshoorn C, et al. Fully conditional specification in multivariate imputation. *J Stat Comput Simul* 2006; **76**: 1049–1064.
13. van Buuren S. Multiple imputation of discrete and continuous data by fully conditional specification. *Stat Meth Med Res* 2007; **16**: 219–242.
14. Bartlett JW, Seaman SR, White IR, et al. Multiple imputation of covariates by fully conditional specification: accommodating the substantive model. *Stat Meth Med Res* 2015; **24**: 462–487.
15. van Buuren S and Groothuis-Oudshoorn K. MICE: multivariate imputation by chained equations in R. *J Stat Softw* 2011; **45**: 1–67.
16. Raghunathan T, Solenberger P and van Hoewyk J. *IVAware: imputation and variance estimation software*, 2016, <http://www.isr.umich.edu/src/smp/ive>
17. Taljaard M, Donner A and Klar N. Imputation strategies for missing continuous outcomes in cluster randomized trials. *Biometr J* 2008; **50**: 329–345.
18. Andridge R. Quantifying the impact of fixed effects modeling of clusters in multiple imputation for cluster randomized trials. *Biometr J* 2011; **53**: 57–74.
19. Drechsler J. Multiple imputation of multilevel missing datarigor versus simplicity. *J Educ Behav Stat* 2015; **40**: 69–95.
20. Schafer J and Yucel R. Computational strategies for multivariate linear mixed-effects models with missing values. *J Comput Graph Stat* 2002; **11**: 437–457.
21. Yucel R. Random covariances and mixed-effects models for imputing multivariate multilevel continuous data. *Stat Model* 2011; **11**: 351–370.
22. Zhao J and Schafer J. *pan: Multiple imputation for multivariate panel or clustered data*. R package version 1.3, 2015.
23. Carpenter J, Goldstein H and Kenward M. REALCOM-IMPUTE software for multilevel multiple imputation with mixed response types. *J Stat Softw* 2011; **45**: 1–14.
24. Quartagno M and Carpenter J. *jomo: a package for multilevel joint modelling multiple imputation*, 2016, R package version 2.2, <http://CRAN.R-project.org/package=jomo>
25. van Buuren S. Multiple imputation of multilevel data. In: Hox J and Roberts J (eds) *The handbook of advanced multilevel analysis*. New York: Taylor & Francis, Ltd, 2010, pp.173–196.
26. van Buuren S. *Flexible imputation of missing data*. Boca Raton: CRC Press, 2012.
27. van Buuren S and Groothuis-Oudshoorn K. *MICE: multivariate imputation by chained equations in R*, 2015, R package version 2.25, <http://CRAN.R-project.org/package=mice>
28. Stewart L and Parmar M. Meta-analysis of the literature or of individual patient data: is there a difference? *Lancet* 1993; **341**: 418–22.
29. Resche-Rigon M, White I, Bartlett J, et al. Multiple imputation for handling systematically missing confounders in meta-analysis of individual participant data. *Stat Med* 2013; **32**: 4890–4905.
30. Jolani S, Debray T, Koffijberg H, et al. Imputation of systematically missing predictors in an individual participant data meta-analysis: a generalized approach using MICE. *Stat Med* 2015; **34**: 1841–1863.
31. West B, Welch K and Galecki A. *Linear mixed models: a practical guide using statistical software*. Boca Raton: CRC Press, 2007.
32. Lee Y, Nelder J and Pawitan Y. *Generalized linear models with random effects: unified analysis via H-likelihood*. Boca Raton: CRC Press, 2006.
33. Laird N and Ware J. Random-effects models for longitudinal data. *Biometrics* 1982; **38**: 963–974.
34. Jackson D, Riley R and White I. Multivariate meta-analysis: potential and promise. *Stat Med* 2011; **30**: 2481–2498.
35. Riley R, Lambert P, Staessen J, et al. Meta-analysis of continuous outcomes combining individual patient data and aggregate data. *Stat Med* 2008; **27**: 1870–1893.

36. Debray TP, Moons KG, Abo-Zaid GMA, et al. Individual participant data meta-analysis for a binary outcome: one-stage or two-stage? *PLoS One* 2013; **8**: e60650.
37. Pinheiro J and Bates D. *Mixed-effects models in S and S-PLUS*. New York: Springer Science & Business Media, 2000.
38. Jackson D, White I and Thompson S. Extending DerSimonian and Laird's methodology to perform multivariate random effects meta-analyses. *Stat Med* 2010; **29**: 1282–1297.
39. Pinheiro J, Bates D, DebRoy S, et al. *nlme: linear and nonlinear mixed effects models*, 2016, R package version 3.1-28, <http://CRAN.R-project.org/package=nlme>
40. R Core Team. *R: a language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria, 2016, <http://www.R-project.org/>
41. Simmonds M, Higgins J, Stewart L, et al. Meta-analysis of individual patient data from randomized trials: a review of methods used in practice. *Clinical Trials* 2005; **2**: 209–217.
42. Thomas D, Radji S and Benedetti A. Systematic review of methods for individual patient data meta-analysis with binary outcomes. *BMC Med Res Meth* 2014; **14**: 79.
43. DerSimonian R and Laird N. Meta-analysis in clinical trials. *Control Clin Trials* 1986; **7**: 177–188.
44. Gasparrini A. *mvmeta: multivariate meta-analysis and meta-regression*, R package version 0.4.5, 2011.
45. Gasparrini A, Armstrong B and Kenward M. Multivariate meta-analysis for non-linear and other multi-parameter associations. *Stat Med* 2012; **31**: 3821–3839.
46. Van Buuren S and Oudshoorn K. Flexible multivariate imputation by MICE. Technical report, Leiden, The Netherlands: TNO Prevention Center, 1999, <http://www.stefvanbuuren.nl/publications/Flexible%20multivariate%20-%20TNO99054%201999.pdf>
47. Bates D, Maechler M, Bolker B, et al. *lme4: linear mixed-effects models using Eigen and S4*, R package version 1.1-12, 2016, <http://CRAN.R-project.org/package=lme4>
48. Mebazaa A, Gayat E, Lassus J, et al. Association between elevated blood glucose and outcome in acute heart failure: results from an international observational cohort. *J Am Coll Cardiol* 2013; **61**: 820–829.
49. Lassus J, Gayat E, Mueller C, et al. Incremental value of biomarkers to clinical variables for mortality prediction in acutely decompensated heart failure: the multinational observational cohort on acute heart failure (MOCA) study. *Int J Cardiol* 2013; **168**: 2186–2194.
50. Riley R, Lambert P and Abo-Zaid G. Meta-analysis of individual participant data: rationale, conduct, and reporting. *Br Med J* 2010; **340**: 521–525.
51. Quartagno M and Carpenter J. Multiple imputation for IPD meta-analysis: allowing for heterogeneity and studies with missing covariates. *Stat Med* 2016; **35**: 2954–2938.
52. Gilks W, Richardson S and Spiegelhalter D. Full conditional distributions. In: *Markov Chain Monte Carlo in practice*. Boca Raton: Springer, 1996, pp.75–88.
53. Liu J, Gelman A, Hill J, et al. On the stationary distribution of iterative imputations. *Biometrika* 2014; **101**: 155–173.
54. Gelman A. Parameterization and Bayesian modeling. *J Am Stat Assoc* 2004; **99**: 537–545.

Appendix I. Conditional multilevel model implied by a bivariate multilevel model

We consider the bivariate multilevel model (3). Let \mathbf{x}_{1i} and \mathbf{x}_{2i} be the n_i -vectors of $\{x_{1ij}\}$'s and $\{x_{2ij}\}$'s, where cluster i has size n_i . We can write equation (3) as

$$\begin{pmatrix} \mathbf{x}_{1i} \\ \mathbf{x}_{2i} \end{pmatrix} \sim N \left(\begin{pmatrix} \mu_1 \mathbf{1} \\ \mu_2 \mathbf{1} \end{pmatrix}, \begin{pmatrix} \Psi_{u11}J + \Sigma_{\varepsilon11}I & \Psi_{u12}J + \Sigma_{\varepsilon12}I \\ \Psi_{u12}J + \Sigma_{\varepsilon12}I & \Psi_{u22}J + \Sigma_{\varepsilon22}I \end{pmatrix} \right)$$

where $\mathbf{1}$ is a n_i -vector of ones, J is a $n_i \times n_i$ matrix of ones, and I is the identity matrix of size n_i . Then the conditional expectation and covariance matrix of \mathbf{x}_{2i} are:

$$E[\mathbf{x}_{2i}|\mathbf{x}_{1i}] = \mu_2 \mathbf{1} + (\Psi_{u12}J + \Sigma_{\varepsilon12}I)(\Psi_{u11}J + \Sigma_{\varepsilon11}I)^{-1}(\mathbf{x}_{1i} - \mu_1 \mathbf{1}), \quad (5)$$

$$\text{Var}[\mathbf{x}_{2i}|\mathbf{x}_{1i}] = (\Psi_{u22}J + \Sigma_{\varepsilon22}I) - (\Psi_{u12}J + \Sigma_{\varepsilon12}I)(\Psi_{u11}J + \Sigma_{\varepsilon11}I)^{-1}(\Psi_{u12}J + \Sigma_{\varepsilon12}I). \quad (6)$$

Lemma 1

For scalars a and b , and identity I and matrix of ones J both of order $n \times n$,

$$(aJ + bI)^{-1} = -\frac{a}{b(an + b)}J + \frac{1}{b}I.$$

Proof

Note $J^2 = nJ$ and $(aJ + bI)(cJ + dI) = (ad + bc + nac)J + bdI$. So $(cJ + dI) = (aJ + bI)^{-1}$ if $bd = 1$ and $ad + bc + nac = 0$. This is true when $d = \frac{1}{b}$ and $c = \frac{-a}{b(b+na)}$.

Using Lemma 1, we obtain the regression coefficient of \mathbf{x}_{2i} on \mathbf{x}_{1i} as

$$(\Psi_{u12}J + \Sigma_{\varepsilon12}I)(\Psi_{u11}J + \Sigma_{\varepsilon11}I)^{-1} = \alpha(n_i)J + \beta I \quad (7)$$

where

$$\alpha(n_i) = \frac{\Psi_{u12}\Sigma_{\varepsilon11} - \Sigma_{\varepsilon12}\Psi_{u11}}{\Sigma_{\varepsilon11}(\Sigma_{\varepsilon11} + n_i\Psi_{u11})} \quad \text{and} \quad \beta = \frac{\Sigma_{\varepsilon12}}{\Sigma_{\varepsilon11}}.$$

Then equation (5) becomes

$$\begin{aligned} E[\mathbf{x}_{2i}|\mathbf{x}_{1i}] &= \mu_2\mathbf{1} + (\alpha(n_i)J + \beta I)(\mathbf{x}_{1i} - \mu_1\mathbf{1}) \\ &= (\mu_2 - \beta\mu_1)\mathbf{1} + \beta\mathbf{x}_{1i} + \alpha(n_i)n_i(\bar{x}_{1i} - \mu_1)\mathbf{1}. \end{aligned} \quad (8)$$

Substituting equation (7) in equation (6) gives

$$\begin{aligned} Var[\mathbf{x}_{2i}|\mathbf{x}_{1i}] &= (\Psi_{u22}J + \Sigma_{\varepsilon22}I) - (\alpha(n_i)J + \beta I)(\Psi_{u12}J + \Sigma_{\varepsilon12}I) \\ &= \gamma(n_i)J + \delta I \end{aligned} \quad (9)$$

where

$$\gamma(n_i) = \Psi_{u22} - \frac{\Sigma_{\varepsilon12}\Psi_{u12}}{\Sigma_{\varepsilon11}} - \frac{(\Psi_{u12}\Sigma_{\varepsilon11} - \Sigma_{\varepsilon12}\Psi_{u11})(\Sigma_{\varepsilon12} + n_i\Psi_{u12})}{\Sigma_{\varepsilon11}(\Sigma_{\varepsilon11} + n_i\Psi_{u11})}$$

and

$$\delta = \Sigma_{\varepsilon22} - \frac{\Sigma_{\varepsilon12}^2}{\Sigma_{\varepsilon11}}.$$

Note that dependence on cluster size disappears as cluster sizes become large, since as $n_i \rightarrow \infty$, $\alpha(n_i)n_i \rightarrow \frac{\Psi_{u12}}{\Psi_{u11}} - \frac{\Sigma_{\varepsilon12}}{\Sigma_{\varepsilon11}}$ and $\gamma(n_i) \rightarrow \Psi_{u22} - \frac{\Sigma_{\varepsilon12}^2}{\Psi_{u11}}$, both of which are constants. Note also that if $\Psi_u = k\Sigma_e$ then $\alpha(n_i) = 0$ and $\gamma(n_i) = \text{constant}$, so again dependence on cluster size disappears.

Appendix 2. Conditional multilevel model implied by a multivariate multilevel model

This appendix generalises the bivariate case of Appendix 1 to consider two sets of p_1 and p_2 variables. Let \mathbf{x}_{kij} be the row-vector of the k th set of variables ($k = 1, 2$) for the j th individual in the i th cluster. We consider the multivariate multilevel model

$$\begin{pmatrix} \mathbf{x}_{1ij} \\ \mathbf{x}_{2ij} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix} + \begin{pmatrix} \mathbf{u}_{1i} \\ \mathbf{u}_{2i} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1ij} \\ \varepsilon_{2ij} \end{pmatrix} \quad (10)$$

where $\begin{pmatrix} \mathbf{u}_{1i} \\ \mathbf{u}_{2i} \end{pmatrix} \stackrel{\text{iid}}{\sim} N\left(\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \Psi_{u11} & \Psi_{u12} \\ \Psi_{u21} & \Psi_{u22} \end{pmatrix}\right)$ and $\begin{pmatrix} \varepsilon_{1i} \\ \varepsilon_{2i} \end{pmatrix} \stackrel{\text{iid}}{\sim} N\left(\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \Sigma_{\varepsilon11} & \Sigma_{\varepsilon12} \\ \Sigma_{\varepsilon21} & \Sigma_{\varepsilon22} \end{pmatrix}\right)$. Here the parameter $\boldsymbol{\mu}_k$ is a $1 \times p_k$ row-vector of means and $\Psi_{ukk'}$ and $\Sigma_{\varepsilon kk'}$ are $p_k \times p_{k'}$ variance and covariance matrices.

Now let \mathbf{X}_{ki} be the $n_i \times p_k$ matrix whose j th row is \mathbf{x}_{kij} . Let $\mathbf{x}_{ki} = \text{vec}(\mathbf{X}_{ki})$ where $\text{vec}(\cdot)$ denotes the operation of stacking the columns. This gives

$$\begin{pmatrix} \mathbf{x}_{1i} \\ \mathbf{x}_{2i} \end{pmatrix} \sim N\left(\begin{pmatrix} \boldsymbol{\mu}_1^T \otimes \mathbf{1} \\ \boldsymbol{\mu}_2^T \otimes \mathbf{1} \end{pmatrix}, \begin{pmatrix} \Psi_{u11} \otimes J + \Sigma_{\varepsilon11} \otimes I & \Psi_{u12} \otimes J + \Sigma_{\varepsilon12} \otimes I \\ \Psi_{u12} \otimes J + \Sigma_{\varepsilon12} \otimes I & \Psi_{u22} \otimes J + \Sigma_{\varepsilon22} \otimes I \end{pmatrix}\right) \quad (11)$$

where $\mathbf{1}$ is a n_i -vector of ones, J is a $n_i \times n_i$ matrix of ones, and I is the identity matrix of size n_i .

Some useful identities for dealing with the Kronecker product $A \otimes B$ are

- (1) For column vectors a and b , $\text{vec}(a^T \otimes b) = a \otimes b$.
- (2) For conformable matrices A , B and X , $(B^T \otimes A)\text{vec}(X) = \text{vec}(AXB)$.
- (3) For conformable matrices A , B , C and D , $(A \otimes B)(C \otimes D) = (AC) \otimes (BD)$.
- (4) For $q \times q$ matrices A and B , and identity I and 1-matrix J both of order $n \times n$,

$$(A \otimes J + B \otimes I)^{-1} = (-(nA + B)^{-1}AB^{-1}) \otimes J + (B^{-1}) \otimes I.$$

We can now express the conditional expectation and covariance matrix of \mathbf{x}_{2i} as:

$$E[\mathbf{x}_{2i}|\mathbf{x}_{1i}] = \boldsymbol{\mu}_2^T \otimes \mathbf{1} + \Lambda(\mathbf{x}_{1i} - \boldsymbol{\mu}_1^T \otimes \mathbf{1}) \quad (12)$$

$$\text{Var}[\mathbf{x}_{2i}|\mathbf{x}_{1i}] = (\Psi_{u22} \otimes J + \Sigma_{\varepsilon 22} \otimes I) - \Lambda(\Psi_{u12} \otimes J + \Sigma_{\varepsilon 12} \otimes I) \quad (13)$$

where $\Lambda = (\Psi_{u12} \otimes J + \Sigma_{\varepsilon 12} \otimes I)(\Psi_{u11} \otimes J + \Sigma_{\varepsilon 11} \otimes I)^{-1}$ is the regression coefficient of \mathbf{x}_{2i} on \mathbf{x}_{1i} . Using identity 4, we obtain $\Lambda = \boldsymbol{\alpha}(n_i) \otimes J + \boldsymbol{\beta} \otimes I$ where

$$\begin{aligned} \boldsymbol{\alpha}(n_i) &= (\Psi_{u21}\Psi_{u11}^{-1}\Sigma_{\varepsilon 11} - \Sigma_{\varepsilon 21})(\Sigma_{\varepsilon 11} + n_i\Psi_{u11})^{-1}\Psi_{u11}\Sigma_{\varepsilon 11}^{-1} \\ \boldsymbol{\beta} &= \Sigma_{\varepsilon 12}\Sigma_{\varepsilon 11}^{-1}. \end{aligned}$$

Then equation (12) becomes

$$E[\mathbf{x}_{2i}|\mathbf{x}_{1i}] = \boldsymbol{\mu}_2^T \otimes \mathbf{1} + (\boldsymbol{\alpha}(n_i) \otimes J + \boldsymbol{\beta} \otimes I)(\mathbf{x}_{1i} - \boldsymbol{\mu}_1^T \otimes \mathbf{1}).$$

Now using the identities above, and recalling $\mathbf{x}_{1i} = \text{vec}(\mathbf{X}_{1i})$, we can show

$$E[\mathbf{X}_{2i}|\mathbf{X}_{1i}] = \boldsymbol{\mu}_2 \otimes \mathbf{1} + J(\mathbf{X}_{1i} - \boldsymbol{\mu}_1 \otimes \mathbf{1})\boldsymbol{\alpha}(n_i)^T + (\mathbf{X}_{1i} - \boldsymbol{\mu}_1 \otimes \mathbf{1})\boldsymbol{\beta}^T.$$

Finally, since $J\mathbf{X}_{1i} = n_i\bar{\mathbf{x}}_{1i} \otimes \mathbf{1}$ where $\bar{\mathbf{x}}_{1i}$ is a row-vector of means, and $J(\boldsymbol{\mu}_1 \otimes \mathbf{1}) = n_i\boldsymbol{\mu}_1 \otimes \mathbf{1}$,

$$E[\mathbf{X}_{2i}|\mathbf{X}_{1i}] = \boldsymbol{\mu}_2 \otimes \mathbf{1} + n_i\{(\bar{\mathbf{x}}_1 - \boldsymbol{\mu}_1) \otimes \mathbf{1}\}\boldsymbol{\alpha}(n_i)^T + \{\mathbf{X}_{1i} - \boldsymbol{\mu}_1 \otimes \mathbf{1}\}\boldsymbol{\beta}^T$$

and

$$E[\mathbf{x}_{2ij}|\mathbf{X}_{1i}] = \boldsymbol{\mu}_2 + n_i(\bar{\mathbf{x}}_1 - \boldsymbol{\mu}_1)\boldsymbol{\alpha}(n_i)^T + (\mathbf{x}_{1ij} - \boldsymbol{\mu}_1)\boldsymbol{\beta}^T. \quad (14)$$

Similarly, equation (13) becomes

$$\begin{aligned} \text{var}(\mathbf{X}_{2i}|\mathbf{X}_{1i}) &= \text{var}(\mathbf{X}_{2i}) - \text{cov}(\mathbf{X}_{2i}, \mathbf{X}_{1i})\text{var}(\mathbf{X}_{1i})\text{cov}(\mathbf{X}_{1i}, \mathbf{X}_{2i}) \\ &= \boldsymbol{\gamma}(n_i) \otimes \mathbf{J} + \boldsymbol{\delta} \otimes \mathbf{I} \end{aligned}$$

where

$$\begin{aligned} \boldsymbol{\gamma}(n_i) &= \Psi_{u22} - \Psi_{u21}\Sigma_{\varepsilon 11}^{-1}\Sigma_{\varepsilon 12} \\ &\quad + (n_i\Psi_{u21} + \Sigma_{\varepsilon 21})(n_i\Psi_{u11} + \Sigma_{\varepsilon 11})^{-1}(\Psi_{u11}\Sigma_{\varepsilon 11}^{-1}\Sigma_{\varepsilon 12} - \Psi_{u12}) \\ \boldsymbol{\delta} &= \Sigma_{\varepsilon 22} - \Sigma_{\varepsilon 21}\Sigma_{\varepsilon 11}^{-1}\Sigma_{\varepsilon 12}. \end{aligned}$$

As in Appendix 1, we note that dependence on cluster size disappears if either (1) $n_i \rightarrow \infty$, since then $n_i\boldsymbol{\alpha}(n_i) \rightarrow \Psi_{u21}\Psi_{u11}^{-1} - \Sigma_{\varepsilon 21}\Sigma_{\varepsilon 11}^{-1}$ and $\boldsymbol{\gamma}(n_i) \rightarrow \Psi_{u22} - \Psi_{u21}\Psi_{u11}^{-1}\Psi_{u12}$, or (2) $\Psi_{u21}\Psi_{u11}^{-1} = \Sigma_{\varepsilon 21}\Sigma_{\varepsilon 11}^{-1}$, since then $\boldsymbol{\alpha}(n_i) = \mathbf{0}$ and $\boldsymbol{\gamma}(n_i) = \Psi_{u22} - \Psi_{u21}\Psi_{u11}^{-1}\Psi_{u12}$ (independent of n_i).

In a MICE procedure, we are interested in all the conditional distributions, not just $[\mathbf{X}_{2i}|\mathbf{X}_{1i}]$. If

$$\begin{pmatrix} \Psi_{u11} & \Psi_{u12} \\ \Psi_{u21} & \Psi_{u22} \end{pmatrix} = k \begin{pmatrix} \Sigma_{\varepsilon 11} & \Sigma_{\varepsilon 12} \\ \Sigma_{\varepsilon 21} & \Sigma_{\varepsilon 22} \end{pmatrix} \text{ then condition (2) above holds for all the conditional distributions.}$$

Appendix 3. Base-case simulations where the missing pattern is either systematically missing or sporadically missing

Tables 8 and 9 show results of the base-case simulation with only sporadically missing variables and only systematically missing variables, respectively. The global percentage of missing data is the same as in the base-case simulations (i.e. 0.44).

Table 8. Base-case simulations with only sporadically missing data. $\Psi_u = 0.5^2 \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}$, $\Sigma_{\varepsilon,1} = \Sigma_{\varepsilon,2} = 0.5$, $\Psi_{\Sigma_{\varepsilon,1}} = \Psi_{\Sigma_{\varepsilon,2}} = 0$.

Method	β_1				$\sqrt{\Psi_{b11}}$	β_2				$\sqrt{\Psi_{b22}}$	Mean run time
	Est mean	Model SE	Emp SE	Cover	Est mean	Est mean	Model SE	Emp SE	Cover	Est mean	
True value	0.5				0.5	1				0.5	
No missing data	0.506	0.113	0.117	0.939	0.491	1.003	0.114	0.109	0.959	0.494	
Complete case	0.506	0.120	0.125	0.943	0.490	1.002	0.120	0.116	0.949	0.491	
Imputation methods, excluding cluster mean from imputation model											
One-stage	0.490	0.105	0.116	0.920	0.422	0.982	0.099	0.114	0.919	0.405	2 min 7 s
Two-stage	0.499	0.120	0.120	0.944	0.500	0.999	0.119	0.114	0.964	0.503	4 min 6 s
Two-stage MM	0.499	0.119	0.119	0.953	0.499	1.000	0.119	0.114	0.962	0.502	1 min 4 s
Imputation methods, including cluster mean in imputation model											
One-stage	0.490	0.105	0.115	0.923	0.421	0.981	0.099	0.114	0.908	0.405	3 min 5 s
Two-stage	0.499	0.119	0.119	0.953	0.499	1.000	0.119	0.114	0.962	0.502	4 min 23 s
Two-stage MM m	0.499	0.119	0.119	0.952	0.499	0.999	0.119	0.113	0.961	0.502	1 min 7 s

Note: Abbreviations as in Table 1.

Table 9. Base-case simulations with only systematically missing data.

Method	β_1				$\sqrt{\Psi_{b11}}$	β_2				$\sqrt{\Psi_{b22}}$	Mean run time
	Est mean	Model SE	Emp SE	Cover	Est mean	Est mean	Model SE	Emp SE	Cover	Est mean	
True value	0.5				0.5	1				0.5	
No missing data	0.495	0.113	0.112	0.951	0.492	1.001	0.113	0.116	0.944	0.490	
Complete case	0.498	0.163	0.166	0.935	0.487	1.008	0.162	0.173	0.916	0.484	
Imputation methods, excluding cluster mean from imputation model											
One-stage	0.483	0.121	0.137	0.916	0.428	0.995	0.113	0.135	0.900	0.445	1 min 32 s
Two-stage	0.480	0.127	0.140	0.922	0.437	1.011	0.121	0.135	0.925	0.458	3 min 10 s
Two-stage MM	0.484	0.125	0.143	0.915	0.434	1.012	0.120	0.133	0.920	0.456	0 min 49 s
Imputation methods, including cluster mean in imputation model											
One-stage	0.481	0.121	0.138	0.912	0.428	0.995	0.114	0.134	0.907	0.445	2 min 49 s
Two-stage	0.476	0.124	0.142	0.911	0.450	1.007	0.120	0.138	0.919	0.459	3 min 26 s
Two-stage MM m	0.479	0.121	0.141	0.908	0.445	1.009	0.120	0.136	0.909	0.460	0 min 51 s

Note: $\Psi_u = 0.5^2 \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}$, $\Sigma_{\varepsilon,1} = \Sigma_{\varepsilon,2} = 0.5$, $\Psi_{\Sigma_{\varepsilon,1}} = \Psi_{\Sigma_{\varepsilon,2}} = 0$. Abbreviations as in Table 1.