

Relative efficiency of joint-model and full-conditional-specification multiple imputation when conditional models are compatible: The general location model

Shaun R Seaman¹ and Rachael A Hughes²

Statistical Methods in Medical Research
2018, Vol. 27(6) 1603–1614

© The Author(s) 2016

Reprints and permissions:

sagepub.co.uk/journalsPermissions.nav

DOI: 10.1177/0962280216665872

journals.sagepub.com/home/smm



Abstract

Estimating the parameters of a regression model of interest is complicated by missing data on the variables in that model. Multiple imputation is commonly used to handle these missing data. Joint model multiple imputation and full-conditional specification multiple imputation are known to yield imputed data with the same asymptotic distribution when the conditional models of full-conditional specification are compatible with that joint model. We show that this asymptotic equivalence of imputation distributions does not imply that joint model multiple imputation and full-conditional specification multiple imputation will also yield asymptotically equally efficient inference about the parameters of the model of interest, nor that they will be equally robust to misspecification of the joint model. When the conditional models used by full-conditional specification multiple imputation are linear, logistic and multinomial regressions, these are compatible with a restricted general location joint model. We show that multiple imputation using the restricted general location joint model can be substantially more asymptotically efficient than full-conditional specification multiple imputation, but this typically requires very strong associations between variables. When associations are weaker, the efficiency gain is small. Moreover, full-conditional specification multiple imputation is shown to be potentially much more robust than joint model multiple imputation using the restricted general location model to misspecification of that model when there is substantial missingness in the outcome variable.

Keywords

Compatibility, chained equations, congeniality, Gibbs sampler, informative margins, linear discriminant analysis, log linear model, missing data

1 Introduction

Estimating the parameters of a regression model of interest (the ‘analysis model’) is often complicated in practice by missing data on the variables in that model. Multiple imputation (MI) is a popular method for dealing with this problem.¹ Values for the missing variables are randomly sampled conditional on the observed variables from distributions thought approximately to describe the association between these variables. The result is an imputed dataset, in which there are no missing data. This imputation is done multiple (say, M) times and the analysis model is fitted separately to each of the resulting M imputed datasets to produce M estimates of the parameters β of this model. Finally, these M estimates are averaged to give an overall estimate of β , known as the ‘Rubin’s Rules (point) estimate’.

MI methods differ in how they randomly sample values for the missing variables. The two most commonly used methods are joint model MI and full-conditional specification (FCS) MI (also known as MI by chained equations).^{2,3} The former involves specifying a joint model for the partially observed variables given the fully observed variables and sampling missing values from their posterior predictive distribution given the observed

¹MRC Biostatistics Unit, Institute of Public Health, Cambridge, UK

²School of Social and Community Medicine, University of Bristol, Bristol, UK

Corresponding author:

Shaun R Seaman, MRC Biostatistics Unit, Institute of Public Health, Forvie Site, Robinson Way, Cambridge CB20SR, UK.

Email: shaun.seaman@mrc-bsu.cam.ac.uk

data. The latter involves specifying a conditional model for each of the partially observed variables given all the other variables and cycling through these models. In special cases, the two approaches are equivalent.⁴ For example, when all the conditional models in FCS MI are linear regressions with main effects and no interactions, FCS MI corresponds to joint model MI using a multivariate normal joint model. Likewise, when all the variables are categorical and the conditional models are saturated logistic regressions, FCS MI is equivalent to joint model MI using a saturated log linear joint model. In general, however, FCS MI is not equivalent to joint model MI.

Liu et al.⁵ (see also Zhu and Raghunathan⁶) showed that, even when FCS MI does not correspond to joint model MI, the distributions from which the two methods sample the missing values (the ‘imputation distributions’) are asymptotically the same when the conditional models used by FCS MI are compatible with a joint model. Compatibility is defined in Section 2. Although this is an important result, the ultimate purpose of MI is to enable the estimation of β , and it is unclear what the consequence of asymptotic equivalence of imputation distributions is for the relative efficiency (RE) of the Rubin’s Rules estimator from FCS MI compared to that from joint model MI. This RE (i.e. the ratio of repeating-sampling variances of the two estimators of β) and, in particular, the asymptotic RE (the ratio as the sample size and M tend to infinity) is the focus of the current article.

When, as is commonly the situation, the partially observed variables consist of both continuous and categorical variables, the conditional models usually employed for them in FCS MI are linear regressions and multinomial logistic regressions, respectively. These are natural choices and are the default options in many statistical packages, e.g. *mice* and *mi* in R, and *ice* and *mi* in STATA. It can be shown that this set of conditional models is compatible with a restricted general location (RGL) joint model. Thus, Liu et al.’s (2014) result implies that FCS MI and joint model MI using the RGL model produce imputations that are asymptotically from the same distribution. Schafer⁷ described how to carry out joint model MI using this RGL model and provided software. So, joint model MI using the RGL model and FCS MI using conditional models compatible with this model are both options for the practicing statistician.

In the current article, we focus on the situation where the aim is to estimate the parameters of the analysis model and MI is used to handle missing data in the variables of that model. We elucidate the relation between joint model MI and FCS MI using compatible conditional models. We focus on the important case where joint model MI uses the RGL model and the (compatible) analysis model is a linear or logistic regression with parameters β . Our goals are (i) to demonstrate that when the RGL model is correctly specified, asymptotic equivalence of imputation distributions does not imply equally asymptotically efficient estimators of β ; (ii) to investigate the magnitude of this difference and how it depends on the strength of associations between outcome and covariates in the analysis model and (iii) to demonstrate that when the joint distribution of the covariates implied by the RGL model is misspecified, FCS MI can be less biased than joint model MI. These goals will be realised using asymptotic calculations and simulation studies.

The structure of the article is as follows. In Section 2, we describe FCS and joint model MI in general and discuss how they are related when the conditional models are compatible. This relation can be one of equivalence in finite samples or asymptotic equivalence. The RGL model is introduced in Section 3. In Section 4, the asymptotic RE of inference from FCS MI with compatible conditional models versus that from the corresponding joint model MI is explored in depth for simple cases of the RGL model: one with a binary and two continuous variables, and one with four binary variables. In addition, the RE of the two MI methods is explored in a more complex situation using data simulated from a realistic data-generating mechanism based on the Barry Caerphilly Growth Study (BCGS).⁸ In Section 5, we discuss and illustrate, using simulated data and data from the National Childhood Development Study (NCDS),⁹ the relative robustness of FCS MI and joint model MI to misspecification of the joint model for the covariates implied by the RGL model. Section 6 contains a discussion.

2 Relation between FCS MI and joint model MI

Let $X = (X_1, \dots, X_K)^\top$ denote a vector of K variables, let $X_{-k} = (X_1, \dots, X_{k-1}, X_{k+1}, \dots, X_K)^\top$ and let $R_k = 1$ if X_k is observed and $R_k = 0$ if X_k is missing. We use subscript i to index the individual in the dataset ($i = 1, \dots, n$). So, $X_i = (X_{i1}, \dots, X_{iK})^\top$, $X_{i,-k}$ and R_{ik} denote the values of X , X_{-k} and R_k for individual i . Let \mathcal{M}_k denote the set of indices of the individuals for whom $R_{ik} = 0$.

In joint model MI, a model $f(X|\theta)$ is specified for the joint distribution of X , with a non-informative prior $p(\theta)$ for the parameters θ in this model. Let Θ denote the parameter space of θ and assume that $p(\theta) > 0 \forall \theta \in \Theta$.

Missing values of \mathbf{X} are imputed from their posterior predictive distribution implied by this model. One way to draw from this distribution is to use the following Gibbs sampler algorithm.⁴ First, replace the missing values by arbitrary starting values. A single iteration of the Gibbs sampler then consists of K steps, in the k th of which the values of $\{X_{ik} : i \in \mathcal{M}_k\}$ are updated. Let $\mathbf{X}_{i,-k}^* = (X_{i1}^*, \dots, X_{i,k-1}^*, X_{i,k+1}^*, \dots, X_{iK}^*)^\top$, where X_{ij}^* equals its observed value X_{ij} if $R_{ij} = 1$ and equals its most recently sampled value if $R_{ij} = 0$. Let $f_k(X_k | \mathbf{X}_{-k}, \boldsymbol{\theta})$ and $f_{-k}(\mathbf{X}_{-k} | \boldsymbol{\theta})$ denote the conditional distributions of X_k given \mathbf{X}_{-k} and the marginal distribution of \mathbf{X}_{-k} , respectively, implied by joint model $f(\mathbf{X} | \boldsymbol{\theta})$. The k th step consists of first sampling $\boldsymbol{\theta}$ from the distribution proportional to $p(\boldsymbol{\theta}) \prod_{i=1}^n f_k(X_{ik} | \mathbf{X}_{i,-k}^*, \boldsymbol{\theta})^{R_{ik}} f_{-k}(\mathbf{X}_{i,-k}^* | \boldsymbol{\theta})$ and then, using this sampled value of $\boldsymbol{\theta}$, sampling X_{ik} from $f_k(X_{ik} | \mathbf{X}_{i,-k}^*, \boldsymbol{\theta})$ for each $i \in \mathcal{M}_k$. These K steps are iterated until the imputed variables converge in distribution.

In FCS MI, a set of K conditional models $\{g_k(X_k | \mathbf{X}_{-k}, \phi_k) : k = 1, \dots, K\}$ is specified for the distribution of each X_k given the remaining variables. Also specified is a non-informative prior $p(\phi_k)$ ($k = 1, \dots, K$) for the parameters ϕ_k in each of these models. Let Φ_k denote the parameter space of ϕ_k and assume that $p_k(\phi_k) > 0 \forall \phi_k \in \Phi_k$. As with the Gibbs sampler, the missing values are first replaced by arbitrary starting values and a single iteration of the FCS algorithm consists of K steps. The k th step involves first sampling ϕ_k from the distribution proportional to $p(\phi_k) \prod_{i=1}^n g_k(X_{ik} | \mathbf{X}_{i,-k}^*, \phi_k)^{R_{ik}}$ and then, using this sampled value of ϕ_k , sampling X_{ik} from $g_k(X_{ik} | \mathbf{X}_{i,-k}^*, \phi_k)$ for each $i \in \mathcal{M}_k$.

Hughes et al.⁴ noted that FCS MI and the Gibbs sampler algorithm (and hence joint model MI) are equivalent when, for each k , the parameters $\boldsymbol{\theta}$ of the joint model can be partitioned (possibly after reparameterisation) into a set of parameters that describe only the conditional distribution of X_k given \mathbf{X}_{-k} and a set of parameters that describe only the marginal distribution of \mathbf{X}_{-k} , and $p(\boldsymbol{\theta})$ implies that these two parameter sets are a priori independent. More formally, for each k ($k = 1, \dots, K$), let $\phi_k = \phi_k(\boldsymbol{\theta})$ and $\phi_{-k} = \phi_{-k}(\boldsymbol{\theta})$ be functions of $\boldsymbol{\theta}$ such that $f_k(X_k | \mathbf{X}_{-k}, \boldsymbol{\theta}) = f_k(X_k | \mathbf{X}_{-k}, \phi_k)$ and $f_{-k}(\mathbf{X}_{-k} | \boldsymbol{\theta}) = f_{-k}(\mathbf{X}_{-k} | \phi_{-k})$. Then joint model MI is equivalent to FCS MI with conditional models $f_k(X_k | \mathbf{X}_{-k}, \phi_k)$ ($k = 1, \dots, K$) if the prior distribution, $p(\phi_k, \phi_{-k})$, of (ϕ_k, ϕ_{-k}) implied by $p(\boldsymbol{\theta})$ can be factorised as $p(\phi_k, \phi_{-k}) = p_k(\phi_k)p_{-k}(\phi_{-k})$ for each k . This ability of the prior to be so factorised has been called the ‘non-informative margins condition’.⁴

The non-informative margins condition cannot hold unless ϕ_k and ϕ_{-k} are distinct parameters, i.e. unless their joint parameter space is the product of their individual parameter spaces. In Section 4.1, we look at two examples where $\boldsymbol{\theta}$ cannot be partitioned into distinct parameters ϕ_k for the conditional distribution and ϕ_{-k} for the marginal distribution. When $\boldsymbol{\theta}$ cannot be partitioned into distinct parameters, data on \mathbf{X}_{-k} indirectly provides information on ϕ_k through the information it provides on ϕ_{-k} . This indirect information is used in the Gibbs sampler but not in FCS MI.

An important theoretical result about the asymptotic relation between FCS and joint model MI was provided by Liu et al.⁵ This result can apply even when the non-informative margins condition is not satisfied. They defined the set of conditional models $\{g_k(X_k | \mathbf{X}_{-k}, \phi_k) : k = 1, \dots, K\}$ to be *compatible* with a joint model $f(\mathbf{X} | \boldsymbol{\theta})$ if (i) for each $\boldsymbol{\theta} \in \Theta$ and for each $k = 1, \dots, K$, there exists a value of $\phi_k \in \Phi_k$ such that $g_k(X_k | \mathbf{X}_{-k}, \phi_k) = f_k(X_k | \mathbf{X}_{-k}, \boldsymbol{\theta})$, and (ii) for each $k = 1, \dots, K$ and for each value of $\phi_k \in \Phi_k$, there exists at least one value of $\boldsymbol{\theta} \in \Theta$ such that $g_k(X_k | \mathbf{X}_{-k}, \phi_k) = f_k(X_k | \mathbf{X}_{-k}, \boldsymbol{\theta})$.

Theorem 1 of Liu et al.⁵ says that if (i) the set of conditional models is compatible with a joint model, (ii) this joint model is correctly specified and (iii) the data are missing at random (MAR), then the total variation distance between the distribution of the imputed data obtained from FCS MI and the distribution of the imputed data obtained from joint model MI tends to zero in probability as the sample size tends to infinity. More informally, we can say that the distribution of the imputed data is asymptotically the same whether one imputes by FCS MI or by joint model MI using the corresponding joint model. Liu et al.⁵ say that ‘iterative imputation [i.e. FCS MI] and joint Bayesian imputation [i.e. joint model MI] are asymptotically the same’ (p. 161).

Three comments are worth making. First, asymptotic equivalence of the imputation distributions of FCS MI and joint model MI does not mean that the two resulting Rubin’s Rules estimators of $\boldsymbol{\beta}$ have the same asymptotic efficiency, as we illustrate in Section 4. Second, when the joint model is misspecified, FCS MI and joint model MI may use different imputation distributions for the missing data, even asymptotically, as we illustrate in Section 5. Third, suppose that \mathbf{X} can be partitioned as $\mathbf{X} = (\mathbf{Z}^\top, \mathbf{X}_A^\top)^\top$, where \mathbf{Z} is fully observed. Then, conditional models for the elements of \mathbf{Z} are not used in the FCS MI algorithm and need not be specified. Likewise, joint model MI requires only a model $f(\mathbf{X}_A | \mathbf{Z}, \boldsymbol{\theta})$ for the conditional distribution of \mathbf{X}_A given \mathbf{Z} ; the marginal distribution of \mathbf{Z} is not used and no model for it need be specified. So, if the conditional models for \mathbf{X}_A are compatible with $f(\mathbf{X}_A | \mathbf{Z}, \boldsymbol{\theta})$, and if $f(\mathbf{X}_A | \mathbf{Z}, \boldsymbol{\theta})$ is correctly specified and the data are MAR, then FCS MI and joint model MI impute missing \mathbf{X}_A from the same distribution asymptotically.

3 The RGL model

Let Y and W be categorical and continuous variables, respectively. A categorical variable with $m > 2$ levels is coded as $m - 1$ indicator variables. The RGL model combines a log linear model with a conditional normal model:

$$P(Y = \mathbf{y}) = \frac{\exp(\boldsymbol{\theta}_y^\top \mathbf{y} + \mathbf{y}^\top \boldsymbol{\theta}_{yy} \mathbf{y})}{\sum_{\mathbf{y}'} \exp(\boldsymbol{\theta}_y^\top \mathbf{y}' + \mathbf{y}'^\top \boldsymbol{\theta}_{yy} \mathbf{y}')} \quad (1)$$

$$W|Y \sim N(\boldsymbol{\theta}_{w0} + \boldsymbol{\theta}_{wy} Y, \boldsymbol{\theta}_v) \quad (2)$$

where $\boldsymbol{\theta}_y$ and $\boldsymbol{\theta}_{w0}$ are parameter vectors and $\boldsymbol{\theta}_{yy}$, $\boldsymbol{\theta}_{wy}$ and $\boldsymbol{\theta}_v$ are parameter matrices. Matrix $\boldsymbol{\theta}_{yy}$ is strictly upper triangular and $\boldsymbol{\theta}_v$ is positive definite. Note that the term $\boldsymbol{\theta}_y^\top \mathbf{y} + \mathbf{y}^\top \boldsymbol{\theta}_{yy} \mathbf{y}$ in equation (1) means that the log linear model includes main effects for Y and all pairwise interactions between pairs of elements of Y . The *mix* library⁷ in R can be used to fit this model and to perform joint model MI based on it.

This RGL model implies that the conditional distribution of any element of W given Y and the remaining elements of W is normal with main effects only. It can also be shown that the RGL model implies that the conditional distribution of any categorical variable in Y given W and the remaining categorical variables has the form of a multinomial logistic regression with main effects only. If this categorical variable is binary, the multinomial logistic regression is just ordinary (binary) logistic regression. Expressions for the log odds ratios (LORs) in this logistic regression in terms of $\boldsymbol{\theta}_y$, $\boldsymbol{\theta}_{yy}$, $\boldsymbol{\theta}_{w0}$, $\boldsymbol{\theta}_{wy}$ and $\boldsymbol{\theta}_v$ are given in Appendix 1. Therefore, if these linear and logistic regressions are used as the conditional models in FCS MI, they are compatible with the RGL joint model. It follows from Theorem 1 of Liu et al.⁵ that if the RGL model is correctly specified and the data are MAR, then FCS MI and joint model MI asymptotically impute from the same distribution.

As mentioned at the end of Section 2, if some elements Z of Y and/or W are fully observed, conditional models are not required for them in FCS MI and they can be conditioned on in joint model MI. Using Y and W now to denote the categorical and continuous variables not included in Z , the resulting joint model is

$$P(Y = \mathbf{y}|Z = \mathbf{z}) = \frac{\exp(\boldsymbol{\theta}_y^\top \mathbf{y} + \mathbf{y}^\top \boldsymbol{\theta}_{yy} \mathbf{y} + \mathbf{y}^\top \boldsymbol{\theta}_{yz} \mathbf{z})}{\sum_{\mathbf{y}'} \exp(\boldsymbol{\theta}_y^\top \mathbf{y}' + \mathbf{y}'^\top \boldsymbol{\theta}_{yy} \mathbf{y}' + \mathbf{y}'^\top \boldsymbol{\theta}_{yz} \mathbf{z})} \quad (3)$$

$$W|Y, Z \sim N(\boldsymbol{\theta}_{w0} + \boldsymbol{\theta}_{wy} Y + \boldsymbol{\theta}_{wz} Z, \boldsymbol{\theta}_v) \quad (4)$$

where $\boldsymbol{\theta}_{yz}$ and $\boldsymbol{\theta}_{yy}$ are strictly upper triangular. We call this the ‘RGL model conditional on Z ’ and write it as ‘CRGL(Z)’. This CRGL(Z) model imposes no constraints on the marginal distribution of Z . Like the RGL model, the CRGL(Z) model implies that the conditional distribution of any categorical variable in Y given W , Z and the remaining categorical variables has the form of a multinomial logistic regression. Expressions for the LORs in this logistic regression are given in Appendix 1. Again, if linear and logistic regressions with main effects only are used as the conditional models in FCS MI, they are compatible with the CRGL(Z) joint model. So, it follows that if the CRGL(Z) model is correctly specified and the data are MAR, then FCS MI and joint model MI using the CRGL(Z) model asymptotically impute from the same distribution. Moreover, since the RGL model implies the CRGL(Z) model, it follows that if the RGL model is correctly specified and the data are MAR, then FCS MI, joint model MI using the CRGL(Z) and joint model MI using the RGL model all asymptotically impute from the same distribution.

Note that, unlike the RGL model, the CRGL(Z) model cannot be fitted using the R *mix* library,⁷ unless Z includes only categorical variables.

Higher order interactions can be added to the log linear models of expressions (1) and (3). The conditional models of FCS MI then require additional interaction terms to remain compatible with this more general RGL or CRGL model. However, we focus on the log linear model with just main effects and pairwise interactions (expressions (1) and (3)) and study the impact of the absence of higher-order terms on the RE of FCS MI and joint model MI for inference about β .

4 Asymptotic RE of RGL versus FCS MI

4.1 Information in the marginal distribution

In Section 2, we noted that when the marginal distribution of X_{-k} contains information about the parameters ϕ_k of the conditional distribution of X_k given X_{-k} , joint model MI uses this information but FCS MI does not. In the

RGL model, when X_k is an element of the vector of continuous variables \mathbf{W} , θ can be partitioned into a priori independent parameters ϕ_k and ϕ_{-k} . So, the marginal distribution of X_{-k} provides no information about ϕ_k .⁴ However, when X_k is one of the categorical variables in \mathbf{Y} , two assumptions of the RGL model make the marginal distribution of X_{-k} informative.⁴

First, expressions (1) and (2) imply that the marginal distribution of \mathbf{W} is a mixture of normal distributions. There is no way to parameterise this marginal distribution more parsimoniously than by using all of $\theta = (\theta_y, \theta_{yy}, \theta_{w0}, \theta_{wy}, \theta_v)$. Therefore, $\phi_{-k} = \theta$.

Second, suppose for simplicity that there are no continuous variables \mathbf{W} , so that the RGL reduces to a log linear model, and that all the categorical variables Y_1, \dots, Y_L are binary. The inclusion of only main effects and pairwise interactions in the log linear model of equation (1) means there are $L(L+1)/2$ parameters. The conditional probability that any one variable, say Y_1 , equals one given the others is the logistic regression form $P(Y_1 = 1 | Y_2, \dots, Y_L) = \text{expit}(\theta_1 + \sum_{j=2}^L \theta_{1j} Y_j)$, where $\theta_{10}, \theta_{12}, \dots, \theta_{1L}$ are parameters. That leaves $L(L-1)/2$ parameters to describe the marginal distribution of (Y_2, \dots, Y_L) . When $L \geq 4$, $L(L-1)/2 < 2^{L-1} - 1$, the number of parameters needed for a saturated model for (Y_2, \dots, Y_L) . This raises the possibility that the marginal distribution of (Y_2, \dots, Y_L) may depend on $\theta_1, \theta_{12}, \dots, \theta_{1L}$, and indeed this is so (see online Appendix). Thus, the marginal distribution contains information about the conditional distribution. This argument extends easily to the general case where $L \geq 4$, categorical variables have more than two categories, and/or there are continuous variables \mathbf{W} .

In the remainder of this section, we study how much this information in the margins affects the asymptotic RE of the Rubin's Rules estimator of β using FCS MI compared to the estimator using joint model MI.

4.2 One binary and two continuous variables

Suppose that data are generated by the RGL model used by Hughes et al.⁴

$$\begin{aligned} Y &\sim \text{Bernoulli}(p) \\ W_1 | Y &\sim \text{Normal}(10 + \gamma_1 Y, 9) \\ W_2 | Y, W_1 &\sim \text{Normal}(9 + 8/9 + W_1/9 + \gamma_2 Y, 8 + 8/9) \end{aligned}$$

where $p=0.1$ or 0.3 , and γ_1 and γ_2 each equal 1, 2, 3 or 4, and W_1 and W_2 are fully observed (Hughes et al.⁴ considered only $p=0.3$ and $\gamma_1 = \gamma_2 = 1$ or $\gamma_1 = \gamma_2 = 3$). This special case of the RGL model with only one binary variable is called the linear discriminant analysis (LDA) model.¹⁰ Using the formula in Appendix 1, it follows that $P(Y = 1 | W_1, W_2) = \text{expit}(\beta_0 + \beta_1 W_1 + \beta_2 W_2)$, where $\beta = (\beta_0, \beta_1, \beta_2)^\top = (\text{logit}(0.3) - 10\gamma_1/9 - 89\gamma_2/80 - \gamma_1^2/18 - 9\gamma_2^2/160, \gamma_1/9 - \gamma_2/80, 9\gamma_2/80)$. Since the standard deviations of W_1 and W_2 are both approximately 3, the LORs β_1 and β_2 are very large when γ_1 and/or γ_2 equals 3 or 4 (e.g. 0.39 and 0.45 when $\gamma_1 = \gamma_2 = 4$). Nevertheless, we include these scenarios in order to investigate what happens in situations of strong associations. More likely scenarios are $\gamma_1 = \gamma_2 = 1$ and $\gamma_1 = \gamma_2 = 2$; (β_1, β_2) is then either (0.099, 0.113) or (0.197, 0.225).

If Y is fully observed, β can be estimated by logistic regression or by LDA. The former makes no assumption about the marginal distribution of $(W_1, W_2)^\top$, whereas LDA assumes that it is a mixture of two normal distributions. LDA is known to be more efficient (in finite samples and asymptotically) than logistic regression when the LDA model is correctly specified, especially when β_1 and β_2 are large or when p is close to 0 or 1.^{10,11}

When Y is partially observed, β can be estimated by using RGL MI or FCS MI and then analysing the imputed data using logistic regression or LDA. Like LDA, RGL MI assumes normality of $(W_1, W_2)^\top$ given Y . Like logistic regression, FCS MI does not assume this. Therefore, if RGL MI and logistic regression analysis are used, the imputer is assuming more than the analyst.¹² If FCS MI and LDA are used, the analyst is assuming more than the imputer. Otherwise, analyst and imputer are making the same assumptions.

For fully observed Y , Table 1 shows the asymptotic RE of LDA compared to logistic regression when $(\gamma_1, \gamma_2) = (1, 1)$, $(2, 2)$ or $(3, 3)$. This was calculated using Monte Carlo integration to evaluate expected information matrices. Results for other (γ_1, γ_2) values are shown in Table 4 of the online Appendix. It is seen that LDA can be more efficient than logistic regression, but that the difference is small unless γ_1 and γ_2 are large and is greater when $p=0.1$ than when $p=0.3$. The largest asymptotic RE when $\gamma_1 \leq 2$ and $\gamma_2 \leq 2$ was 104%, although it did rise to 142% when $\gamma_1 = \gamma_2 = 4$ and $p=0.1$.

These complete-data results suggest RGL MI may often not be much more efficient than FCS MI when Y is partially observed. To investigate this, we assumed that Y is missing with probability 0.5, either completely at

Table 1. Percentage asymptotic REs of LDA versus logistic regression analysis when using complete data.

p	γ_1	γ_2	Regression coefficient		
			Intercept	W_1	W_2
0.1	1	1	99.4	100.4	100.3
0.1	2	2	103.7	103.3	103.6
0.1	3	3	116.6	112.7	114.0
0.3	1	1	99.2	100.2	100.1
0.3	2	2	101.7	101.7	101.8
0.3	3	3	109.5	107.0	107.6

Table 2. Percentage asymptotic REs of RGL MI versus FCS MI for four different analysis models.

		Analysis and regression coefficient									
γ_1	γ_2	E(Y)	lr(0)	lr(W_1)	lr(W_2)	ld(0)	ld(W_1)	ld(W_2)	ln(0)	ln(Y)	ln(W_1)
<i>p</i> = 0.1 and MCAR											
1	1	99.9	100.4	100.3	100.4	100.0	100.1	100.1	100.0	100.1	100.0
2	2	99.9	106.6	104.5	105.1	101.8	101.3	101.5	100.1	101.5	100.3
3	3	100.1	123.9	117.3	119.1	105.8	104.5	105.0	100.9	104.5	101.5
<i>p</i> = 0.1 and MAR											
1	1	99.8	100.5	100.1	100.3	100.0	99.8	100.0	100.0	100.0	100.0
2	2	102.2	110.1	108.4	106.3	103.7	103.7	102.1	100.5	102.0	101.0
3	3	107.6	138.2	134.1	125.4	111.8	113.6	108.1	103.8	107.2	105.8
<i>p</i> = 0.3 and MCAR											
1	1	100.0	100.2	100.2	100.1	100.0	100.1	100.0	100.0	100.0	100.0
2	2	100.1	103.7	102.4	102.6	101.1	100.8	100.7	100.1	100.7	100.2
3	3	100.6	114.6	109.9	110.8	104.0	102.9	103.0	100.9	102.7	101.3
<i>p</i> = 0.3 and MAR											
1	1	100.0	100.2	100.2	100.1	100.0	100.0	99.9	100.0	99.9	100.0
2	2	100.6	104.2	103.2	103.0	101.4	101.3	101.0	100.2	100.9	100.4
3	3	103.2	117.6	114.5	113.3	105.8	106.1	104.5	101.7	103.9	102.9

E(Y): marginal mean of Y; lr: logistic regression; ld: linear discriminant analysis; ln: linear regression; (V): the regression coefficient associated with variable V (with (0) meaning the intercept); MAR: missing at random; MCAR: missing completely at random.

random or at random with probability $P(R = 0|W_1, W_2) = \text{expit}(c - W_1/3)$ (*c* was chosen to give $P(R = 1) = 0.5$). Using the formula in Theorem 1 of Robins and Wang,¹³ we calculated the asymptotic REs of RGL MI versus FCS MI for the Rubin’s Rules estimators. This is the RE for an infinite sample size and $M = \infty$ imputations. Monte Carlo integration was used to evaluate the expectations in the Robins and Wang formula. We considered four analyses: logistic regression, LDA, linear regression of W_2 on Y and W_1 , and estimation of the marginal mean of Y .

Table 2 shows results for $(\gamma_1, \gamma_2) = (1, 1), (2, 2)$ and $(3, 3)$. Results for other (γ_1, γ_2) values are in Table 5 in the online Appendix. As expected, RGL MI is only slightly more efficient than FCS MI unless γ_1 and/or γ_2 are large, and the efficiency gain is greater for $p = 0.1$ than for $p = 0.3$. Efficiency gains are much greater for the logistic regression analysis than for the other three analyses but are still $\leq 10\%$ unless $\gamma_1 > 2$ or $\gamma_2 > 2$.

It is interesting that FCS MI is less efficient than RGL MI when the analysis is LDA. This shows that, for the RGL model, asymptotic efficiency is lost by the imputer assuming less than the analyst. Meng¹² gave a different example of an imputer assuming less than an analyst and showed there was no loss of asymptotic efficiency in that case. It is also interesting that RGL MI is more efficient than FCS MI when the analysis is logistic regression. This is an example of the imputer assuming more than the analyst. Since FCS MI with an infinite number of imputations followed by logistic regression analysis is asymptotically equivalent to logistic regression analysis using only complete cases (because individuals with missing outcome provide no information in logistic regression when all covariates are fully observed), the greater efficiency of RGL MI followed by logistic regression analysis is an illustration of ‘super-efficiency’.^{12,14}

Next, we investigated whether the asymptotic REs in Table 1 reflect the REs in finite samples. Table 6 in the online Appendix shows, for two scenarios, the RE of LDA using complete data versus logistic regression using complete data for a variety of sample sizes. The REs were estimated using 10,000 simulated datasets. Table 7 in the online Appendix shows, for the same two scenarios (and using the same 10,000 simulated datasets), the finite-sample RE of RGL MI versus FCS MI (using $M = 50$ imputations) for the four analyses of Table 2. The finite-sample REs are similar to the asymptotic REs. Note that, as expected, the Rubin's Rules point estimators were approximately unbiased for all methods (data not shown).

4.3 Four binary variables

Now suppose that data are generated by the log linear model with $L = 4$ binary variables, $P(Y_1, Y_2, Y_3, Y_4) \propto \exp\left(\sum_{j=1}^4 \theta_j Y_j + \sum_{j=1}^3 \sum_{k=2}^4 \theta_{jk} Y_j Y_k\right)$, where $\theta_j = -0.5$ for $j = 2, 3, 4$; $\theta_{jk} = 0.5$ for $(j, k) = (2, 3), (2, 4), (3, 4)$; $(\theta_{12}, \theta_{13}, \theta_{14}) = (0.33, 0.67, 1.00), (0.67, 1.33, 2.00), (1, 2, 3)$ or $(3, 3, 3)$; and θ_1 is chosen to make $P(Y_1 = 1) = 0.3$. The parameters $\boldsymbol{\beta} = (\theta_1, \theta_{12}, \theta_{13}, \theta_{14})^\top$ can be estimated by fitting either this log linear model or the logistic regression model $P(Y_1 = 1 | Y_2, Y_3, Y_4) = \text{expit}\left(\theta_1 + \sum_{j=2}^4 \theta_{1j} Y_j\right)$ that it implies. We calculated the RE of these two methods when the data were complete. We also calculated the RE of RGL MI versus FCS MI when Y_1 was missing with probability 0.5 completely at random or at random and analysis was either by logistic regression or by fitting the log linear model.

Detailed results are given in the online Appendix. In summary, we found that analysing the complete data by fitting the log linear model was hardly any more efficient than using logistic regression: all REs were less than 107%. Likewise, analysing RGL MI was not much more efficient than FCS MI: all REs were less than 116% when analysis was by logistic regression and all were less than 108% when analysis was by fitting the log linear model. These maximum REs required strong associations between variables, i.e. $(\theta_{12}, \theta_{13}, \theta_{14}) = (3, 3, 3)$. With more moderate associations, REs did not exceed 105%. It appears therefore that the marginal distribution of (Y_2, Y_3, Y_4) contains little information about $\theta_1, \theta_{12}, \theta_{13}$ and θ_{14} . We also considered data-generating mechanisms with $L = 6$ variables or changed the parameters of interest to $\boldsymbol{\beta} = (\theta_4, \theta_{14}, \theta_{24}, \theta_{34})^\top$, i.e. parameters of the regression of Y_4 on Y_1, Y_2 and Y_3 , so that the partially observed variable (Y_1) was a covariate. In all cases, REs were less than 108% (data not shown).

4.4 Simulation study based on BCGS

To investigate RE of RGL MI versus FCS MI in a realistic setting, we carried out a simulation study based on real data from the BCGS. The BCGS is a follow-up study of a dietary intervention randomised controlled trial of pregnant women and their offspring.⁸ Participants in the original trial were followed up until offspring were five years old. When aged 25, these offspring were invited to participate in a follow-up study. There were 951 offspring in the trial, of whom 712 participated in the follow-up study.

For the simulation study, we considered eight variables: sex, childhood weight (at age 5), adult overweight (a binary indicator of $\text{BMI} \geq 25$), ex-smoker and height (all at age 25), father's and mother's weights and father's social class. We considered as our analysis of interest a logistic regression of adult overweight on the other variables. Adult overweight and adult height were missing on (the same) 272 of the 951 offspring; ex-smoker and father's weight were missing on, respectively, 241 and 149 offspring. Sex was fully observed, and there were a total of 12 missing values on the remaining three variables. Among the 679 offspring with observed outcome, there were only 109 missing values, 98 of which were for father's weight.

Simulated datasets were created as follows. First, we fitted the RGL model of equations (1) and (2) to the BCGS data. Then, the fitted model was used to generate complete data on the eight variables for each of 951 hypothetical individuals independently. Missingness was then imposed using missingness models whose parameters were estimated from the BCGS data.

The simulation study was in two parts. In Part I, three of the continuous variables (father's, mother's and childhood weights) were treated as auxiliary variables, i.e. they were included in the imputation model but not in the analysis model; the other variables were included in both models. Using auxiliary variables in the imputation model may increase efficiency and make the MAR assumption more plausible.¹⁵ With auxiliary variables, it may be worth imputing a missing outcome even if the covariates in the analysis model are fully observed.

In Part II, there were no auxiliary variables: all eight variables were included in the analysis model. In the absence of auxiliary variables, Von Hippel¹⁵ recommended including all individuals in the imputation step but

then excluding those with imputed outcomes before fitting the analysis model to the imputed data, in order to reduce bias caused by a possibly misspecified imputation model. This approach is valid when (i) the data are MAR, (ii) the model for the conditional distribution of outcome given covariates implied by the imputation model is the same as the analysis model and (iii) the analysis model is correctly specified. We therefore analysed imputed data both before and after excluding imputed outcomes. As the proportion of missing covariate values among offspring with observed outcome was small in the BCGS dataset, we increased this proportion for the simulation (Part II only).

For both parts, we checked that the RGL model was not an obvious poor fit to the BCGS data by comparing the LORs from a complete-case logistic regression of adult overweight on the other variables with the corresponding LORs implied by the fitted RGL model. The estimates were similar, providing some reassurance.

We considered several simulation scenarios, by varying the strength of association between the auxiliary variables and outcome (in Part I) and the amount of missingness in the covariates (in Part II). For each scenario, we simulated 1000 datasets using R. RGL MI was performed using the *mix* library in R; FCS MI used the *ice* package in STATA; $M=100$ imputations were used. Full details are given in the online Appendix.

To summarise the results, in Part I the maximum RE of RGL MI versus FCS MI was 104% (this was for LOR of ex-smoker). In Part II, the covariate with the highest RE was ex-smoker; this RE was 105% when 33% of ex-smoker values were missing and 111% when 54% of ex-smoker values were missing. When imputed outcomes were excluded before fitting the analysis model, these maximum REs decreased to, respectively, 102% and 108%. Full results are in the online Appendix.

5 Robustness of FCS and joint model MI

In Section 4, we demonstrated that RGL MI can be more efficient than FCS MI but that the gains seem to be small unless associations between variables are very strong. In this section, we show that these efficiency gains can come at the price of bias when the RGL model is misspecified. In Section 5.1, we modify the RGL model used in Section 4.2 so that W_1 is not normally distributed given Y . It is now a CRGL(W_1) model but not a RGL model. We show that when W_1 is fully observed and Y is partially observed, logistic regression gives unbiased estimation when imputation is by FCS MI but not when RGL MI is used. In Section 5.2, we modify the log linear model of Section 4.3 by introducing a third-order interaction between Y_2 , Y_3 and Y_4 . The log linear model with only main effects and pairwise interactions is now misspecified. We show that this causes bias when RGL MI is used but not when FCS MI is used, because FCS MI makes no assumption about the distribution of fully observed variables. In the online Appendix, we present a realistic analysis of data from the NCDS,⁹ which illustrates that use of RGL MI can lead to serious bias in a situation where FCS MI does not.

5.1 One binary and two continuous variables

We simulated data from the following modification of the RGL model of Section 4.2.

$$W_1 \sim \text{Gamma}(2, 2) \quad (5)$$

$$Y|W_1 \sim \text{Bernoulli}(\text{expit}(-1.9 + W_1)) \quad (6)$$

$$W_2|Y, W_1 \sim \text{Normal}(10 + \gamma Y + W_1, 9) \quad (7)$$

Now, W_1 is no longer normally distributed given Y . This is a CRGL(W_1) model but not a RGL model. We considered eight scenarios defined by the value of γ (1 or 3), by which variables were partially observed (either just Y or both Y and W_2), and by whether data were missing completely at random (MCAR) or MAR. The probability that each partially observed variable was observed was 0.5 if data were MCAR and $\text{expit}(-1 + W_1)$ if MAR (this gives a marginal probability of missingness of 0.5). In scenarios where both Y and W_2 were missing, their missingness was independent given W_1 . For each scenario, we generated 1000 datasets each of size $n=1000$.

Missing data were imputed using either FCS MI or RGL MI, with $M=50$ imputations. Four analyses were carried out: estimating $E(Y)$ by the sample mean of Y ; estimating parameters β_0 , β_1 and β_2 of $P(Y=1|W_1, W_2) = \text{expit}(\beta_0 + \beta_1 W_1 + \beta_2 W_2)$ using either logistic regression or LDA; and estimating the parameters of the linear regression of W_2 on Y and W_1 . Using the formula in Appendix 1, $(\beta_0, \beta_1, \beta_2) = (-3.067, 0.889, 0.111)$ when $\gamma=1$ and $(-5.733, 0.667, 0.333)$ when $\gamma=3$.

Since W_1 is fully observed and the CRGL(W_1) model is correctly specified, logistic regression analysis of data imputed by FCS MI should yield asymptotically unbiased estimators, whereas imputing using

Table 3. Mean estimates when RGL model is misspecified and $\gamma = 1$.

	Analysis and regression coefficient									
	E(Y)	lr(0)	lr(W ₁)	lr(W ₂)	ld(0)	ld(W ₁)	ld(W ₂)	ln(0)	ln(Y)	ln(W ₁)
True	0.300	-3.067	0.889	0.111	-3.067	0.889	0.111	10.000	1.000	1.000
cdata	0.303	-3.091	0.890	0.113	-3.232	0.987	0.113	10.001	1.012	0.995
Y MCAR, and W ₁ and W ₂ fully observed										
FCS	0.304	-3.101	0.895	0.114	-3.241	0.992	0.114	9.999	1.011	0.996
RGL	0.296	-3.276	1.009	0.114	-3.456	1.131	0.114	10.024	1.013	0.979
Both Y and W ₂ MCAR, and W ₁ fully observed										
FCS	0.304	-3.105	0.900	0.113	-3.245	0.997	0.113	10.006	0.998	0.993
RGL	0.296	-3.287	1.014	0.114	-3.468	1.136	0.114	10.032	1.004	0.974
All of Y, W ₁ and W ₂ MCAR										
FCS	0.304	-3.124	0.916	0.113	-3.255	1.004	0.113	10.003	0.996	0.996
RGL	0.296	-3.299	1.028	0.114	-3.466	1.140	0.114	10.028	1.001	0.978
Y MAR, and W ₁ and W ₂ fully observed										
FCS	0.303	-3.102	0.891	0.114	-3.237	0.983	0.114	9.998	1.012	0.998
RGL	0.303	-3.271	1.047	0.114	-3.445	1.161	0.114	10.024	1.012	0.972
Both Y and W ₂ MAR, and W ₁ fully observed										
FCS	0.304	-3.111	0.898	0.114	-3.247	0.989	0.114	10.005	1.004	0.993
RGL	0.303	-3.290	1.051	0.115	-3.465	1.166	0.115	10.027	1.010	0.971

lr: logistic regression; ld: linear discriminant analysis; ln: linear regression; (V): the regression coefficient associated with variable V (with (0) meaning the intercept); MCAR: missing completely at random; FCS: full-conditional specification; RGL: restricted general location; MAR: missing at random.

RGL MI or analysing using LDA may yield asymptotically biased estimators, because the RGL model is misspecified. Table 3 shows the means of the parameter estimates when $\gamma = 1$. Results for $\gamma = 3$ are given in Table 8 of the online Appendix. It can be seen that LDA gives a biased estimate of the LOR of W_1 whether one uses complete data, FCS MI or RGL MI. Provided that the analysis is logistic regression or linear regression, there is no bias when using complete data or FCS MI. For RGL MI, on the other hand, there is bias in the coefficient of W_1 in the logistic regression and linear regression analyses. These biases are small for linear regression and are slightly greater for $\gamma = 3$ than for $\gamma = 1$.

For the scenarios where Y and W_2 were both partially observed, we also applied logistic regression to the datasets imputed by RGL MI after excluding the individuals whose Y value had been imputed. Similarly, we applied linear regression to the imputed datasets after excluding the individuals whose W_2 value had been imputed. This strategy of excluding imputed outcomes before analysing the data has been advocated by Von Hippel¹⁵ as a way of reducing bias caused by a possibly misspecified imputation model. Table 9 in the online Appendix shows the results. Most or all of the bias has been removed for logistic regression but none has been removed for linear regression. Note that Von Hippel did not recommend this approach when MI is done using strong auxiliary variables.

Finally, in scenarios where Y_1 and W_2 were both MCAR, we additionally imposed 10% missingness on W_1 . Tables 3 and 8 show that although there is some bias for logistic regression analysis when FCS MI is used, this is much less than when RGL MI is used.

5.2 Four binary variables

Consider again the data generating mechanism of Section 4.3 but now suppose that the log linear model contains an additional third-order interaction:

$$\log P(Y_1, Y_2, Y_3, Y_4) \propto \exp\left(\sum_{j=1}^4 \theta_j Y_j + \sum_{j=1}^3 \sum_{k=2}^4 \theta_{jk} Y_j Y_k - 2Y_2 Y_3 Y_4\right)$$

The log linear model with only main effects and pairwise interactions is now misspecified. Table 10 in the online Appendix shows, for four true values of $\beta = (\theta_{12}, \theta_{13}, \theta_{14})^T$, the mean estimates of β when the complete data or imputed data are analysed by logistic regression or by fitting the log linear model. This shows that there is bias

when fitting the log linear model (even to the complete data) or when imputation is by RGL MI, and that there is no bias when the complete data or data imputed by FCS MI are analysed by logistic regression. Note that, unlike the normality assumption of the RGL model, which is an intrinsic feature of that model, higher-order interactions can be allowed in the RGL model, but in practice, this might not be done.

6 Discussion

FCS and joint model MI yield imputed data with the same asymptotic distribution when the conditional models used by FCS MI are compatible with the joint model. However, we have shown that this asymptotic equivalence in terms of the imputation distribution does not imply that FCS and joint model MI yield equally asymptotically efficient estimates of the parameters in the analysis model. Moreover, FCS MI can be more robust than joint model MI to misspecification of the joint model. We focussed on the RGL model. The efficiency gain from using joint model MI with this model (RGL MI) rather than the corresponding FCS MI appears to be small, except when the outcome is categorical and has a large proportion of missingness and very strong associations exist between the outcome and covariates. On the other hand, we have shown that if the RGL model is misspecified, RGL MI can be much more biased than FCS MI in this same situation, even when covariate-outcome associations are weaker.

Robustness of RGL MI can be improved by including additional interactions in the model (this could have been done in, e.g. the analysis of the NCDS data in the online Appendix) or by conditioning on fully observed variables Z (the CRGL(Z) model). However, the R *mix* library cannot be used to fit the CRGL(Z) model or to carry out joint model MI with this model, unless Z includes only categorical variables. Bayesian modelling software, such as WinBUGS, could be used, but this requires more specialist programming skills. Robustness of RGL MI can also be improved by excluding individuals with imputed outcomes from the analysis. This approach was advocated by Von Hippel (2007),¹⁵ at least when there are no strong auxiliary variables. In the absence of auxiliary variables and when data are MAR, excluding individuals with imputed outcomes causes no loss of efficiency when the analysis is by linear regression or LDA, and causes no bias and is likely to cause little loss of efficiency when analysis is by logistic regression (especially when those with missing outcomes also have missing values in covariates). Conversely, Sullivan et al.¹⁶ show that excluding imputed outcomes can cause significant bias when auxiliary variables are strongly associated with both the outcome and missingness in that outcome. They did not, however, investigate situations where the imputation model is misspecified.

Although careful assessment of goodness of fit of an imputation model could detect poor fit of that model, we suspect that this may often not be done in practice. For this reason, FCS MI may be safer than RGL MI when a large proportion of outcomes are missing, unless imputed outcomes are excluded from the subsequent analysis. Since, as Sullivan et al.¹⁶ noted, this exclusion can itself induce another bias, we suggest that a good approach may be to use FCS MI imputing the outcome last and including imputed outcomes in the analysis. Our results indicate that the efficiency loss from using FCS rather than joint model MI is unlikely to be significant in practice.

Several comparisons of joint model MI with FCS MI have been published (e.g. Lee and Carlin¹⁷ and Kropko et al.¹⁸ and references therein). However, these have focussed on joint model MI using a multivariate normal model. The distributions of the imputed data from this joint normal model MI and from FCS MI are not asymptotically equivalent, unless all variables are continuous. These published comparisons have generally noted little difference in efficiency, and relative robustness depended on how categorical variables were handled by joint normal model MI.

An alternative to RGL (or CRGL) MI and FCS MI is joint model MI under the latent normal model of Goldstein and Carpenter.^{2,19} This can be implemented using REALCOM-IMPUTE or the jomo package in R. This software allows conditioning on fully observed variables. This approach also extends to multi-level data by using random effects. Unlike the RGL model, the latent normal model does not imply conditional distributions that are linear or logistic/multinomial regressions. This is why we compared FCS MI with RGL MI rather than with joint model MI under the latent normal model. It also means that, in general, there is incompatibility (sometimes known as ‘uncongeniality’¹²) between a latent normal imputation model and a linear or logistic regression analysis model. Nevertheless, while some forms of incompatibility (uncongeniality) between imputation and analysis models (e.g. an imputation model that ignores an interaction present in the analysis model¹²) may cause substantial bias in the estimates of the parameters of the analysis model, other forms may often not matter in practice.⁷ Moreover, MI under the latent normal model may be more robust than RGL MI to model misspecification; more research on this is needed.

A limitation of our work is that, because it was not feasible to study every possible data-generating mechanism, we cannot rule out the possibility that there are scenarios in which large efficiency gains are possible without

requiring strong associations between variables, although this seems unlikely. We focussed on parameter estimation. It is plausible that many of our conclusions could apply when the model of interest is used for prediction, since in that case the linear predictor is a weighted average of the individual parameters. However, further research is warranted into the RE and relative robustness of joint model MI and FCS MI when the ultimate aim is prediction, classification or clustering. Another direction of future research would be to compare FCS and RGL MI when data are missing not at random, the CRGL model is misspecified or FCS MI does not use compatible conditional models.²⁰ It is possible that FCS MI using incompatible conditional models may be more efficient than joint model MI using a misspecified joint model, especially when those conditional models have been chosen to fit well to the observed data.

In conclusion, FCS MI may be preferable to joint model MI using the compatible joint model, viz. the RGL model: it is more robust and is usually only slightly less efficient.

Acknowledgements

We thank Prof Yoav Ben-Shlomo and Dr Anne McCarthy, from the School of Social and Community Medicine, University of Bristol, for granting access to the BCGS data; Prof Chris Power for assistance in obtaining the NCDS data, and the Centre for Longitudinal Studies for providing these data; and Drs Finbarr Leacy, Ian White and Jonathan Bartlett for helpful discussions.

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: SRS and RAH were supported by Medical Research Council [grant numbers U105260558, MR/J013773/1].

References

1. Little RJA and Rubin DB. *Statistical analysis with missing data*, 2nd ed. Hoboken, NJ: Wiley, 2002.
2. Carpenter J and Kenward M. *Multiple imputation and its applications*. Hoboken, NJ: Wiley, 2013.
3. van Buuren S. *Flexible imputation of missing data*. Boca Raton, FL: Chapman & Hall/CRC, 2012.
4. Hughes RA, White IR, Seaman SR, et al. Joint modelling rationale for chained equations. *BMC Med Res Methodol* 2014; **14**: 28.
5. Liu J, Gelman A, Hill J, et al. On the stationary distribution of iterative imputations. *Biometrika* 2014; **101**: 155–173.
6. Zhu J and Raghunathan T. Convergence properties of a sequential regression multiple imputation algorithm. *J Am Stat Assoc* 2015; **110**: 1112–1124.
7. Schafer JL. *Analysis of incomplete multivariate data*. Boca Raton, FL: Chapman & Hall, 1997.
8. McCarthy A, Hughes R, Tilling K, et al. Birth weight; postnatal, infant, and childhood growth; and obesity in young adulthood: evidence from the Barry Caerphilly growth study. *Am J Clin Nutr* 2007; **86**: 907–913.
9. Thomas C, Hypponen E and Power C. Prenatal exposures and glucose metabolism in adulthood. *Diabetes Care* 2007; **30**: 918–924.
10. Efron B. The efficiency of logistic regression compared to normal discriminant analysis. *J Am Stat Assoc* 1975; **70**: 892–898.
11. Bull SB and Donner A. Derivation of large sample efficiency of multinomial logistic regression compared to multiple group discriminant analysis. In: MacNeill IB, Umphrey GJ, Donner A, et al. (eds) *Biostatistics* (volume 38 of The University of Western Ontario Series in philosophy of science). Dordrecht: Springer Netherlands, 1987, pp. 177–197.
12. Meng X-L. Multiple-imputation inferences with uncongenial sources of input. *Stat Sci* 1994; **9**: 538–573.
13. Robins J and Wang N. Inference for imputation estimators. *Biometrika* 2000; **87**: 113–24.
14. Rubin DB. Multiple imputation after 18 + years. *J Am Stat Assoc* 1996; **91**: 473–489.
15. Von Hippel P. Regression with missing Ys: an improved strategy for analysing multiply imputed data. *Soc Methodol* 2007; **37**: 83–117.
16. Sullivan TR, Salter AB, Ryan P, et al. Bias and precision of the ‘multiple imputation, then deletion’ method for dealing with missing outcome data. *Am J Epidemiol* 2015; **182**: 528–534.
17. Lee KJ and Carlin JB. Multiple imputation for missing data: fully conditional specification versus multivariate normal imputation. *Am J Epidemiol* 2010; **171**: 624–632.

18. Kropko J, Goodrich R, Gelman A, et al. Multiple imputation for continuous and categorical data: comparing joint multivariate normal and conditional approaches. *Pol Anal* 2014; **22**: 497–519.
19. Goldstein H, Carpenter J, Kenward MG, et al. Multilevel models with multivariate mixed response types. *Stat Model* 2009; **9**: 173–197.
20. Li F, Yu Y and Rubin DB. Imputing missing data by fully conditional models: some cautionary examples and guidelines. Technical report, Duke University, <http://www2.stat.duke.edu/fl35/publications.html> (2012, accessed 10 August 2016).
21. Power C and Elliott J. Cohort profile: 1958 British birth cohort (National Child Development Study). *Int J Epidemiol* 2006; **35**: 34–41.

Appendix I

LORs implied by CGRL model

Consider the CRGL(\mathbf{Z}) model of expressions (3) and (4). Partition \mathbf{Y} as $\mathbf{Y} = (Y_A, \mathbf{Y}_B^\top)^\top$, where Y_A is a binary variable (and not a dummy variable for a categorical variable with more than two levels). Similarly, partition $\boldsymbol{\theta}_y$ as $\boldsymbol{\theta}_y = (\boldsymbol{\theta}_{yA}, \boldsymbol{\theta}_{yB}^\top)^\top$ and similarly with $\boldsymbol{\theta}_{yy}$, $\boldsymbol{\theta}_{yz}$, $\boldsymbol{\theta}_{w0}$, $\boldsymbol{\theta}_{wy}$ and $\boldsymbol{\theta}_{wz}$.

Equation (3) implies that

$$P(Y_A = 1 | \mathbf{Y}_B, \mathbf{Z}) = \frac{\exp(\boldsymbol{\theta}_{yA} + \boldsymbol{\theta}_{yAyB}^\top \mathbf{Y}_B + \boldsymbol{\theta}_{yAz}^\top \mathbf{Z})}{1 + \exp(\boldsymbol{\theta}_{yA} + \boldsymbol{\theta}_{yAyB}^\top \mathbf{Y}_B + \boldsymbol{\theta}_{yAz}^\top \mathbf{Z})}$$

By using Bayes' Theorem, it is easy to show that

$$P(Y_A = 1 | \mathbf{Y}_B, \mathbf{Z}, \mathbf{W}) = \frac{\exp(\beta_0 + \boldsymbol{\beta}_1^\top \mathbf{Y}_B + \boldsymbol{\beta}_2^\top \mathbf{Z} + \boldsymbol{\beta}_3^\top \mathbf{W})}{1 + \exp(\beta_0 + \boldsymbol{\beta}_1^\top \mathbf{Y}_B + \boldsymbol{\beta}_2^\top \mathbf{Z} + \boldsymbol{\beta}_3^\top \mathbf{W})}$$

where

$$\begin{aligned}\boldsymbol{\beta}_3 &= \boldsymbol{\theta}_v^{-1} \boldsymbol{\theta}_{wyA} \\ \beta_0 &= \boldsymbol{\theta}_{yA} - \boldsymbol{\theta}_{w0}^\top \boldsymbol{\beta}_3 - 0.5 \boldsymbol{\theta}_{wyA}^\top \boldsymbol{\beta}_3 \\ \boldsymbol{\beta}_1 &= \boldsymbol{\theta}_{yAyB} - \boldsymbol{\theta}_{wyB}^\top \boldsymbol{\beta}_3 \\ \boldsymbol{\beta}_2 &= \boldsymbol{\theta}_{yAz} - \boldsymbol{\theta}_{wz}^\top \boldsymbol{\beta}_3\end{aligned}$$

The CRGL(\mathbf{Z}) model reduces to the RGL model when \mathbf{Z} is empty. Hence, the above expressions for β_0 , $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_3$ apply to the RGL model if $\boldsymbol{\beta}_2$, $\boldsymbol{\theta}_{yAz}$ and $\boldsymbol{\theta}_{wz}$ are omitted.