Published in final edited form as:

Stat Med. 2017 September 30; 36(22): 3507–3532. doi:10.1002/sim.7388.

A comparison of existing methods for multiple imputation in individual participant data meta-analysis

Deborah Kunkel and Eloise E. Kaizar*

Department of Statistics, Ohio State University, 1958 Neil Avenue, Cockins Hall, Room 404, Columbus, OH 43210-1247

Abstract

Multiple imputation is a popular method for addressing missing data, but its implementation is difficult when data have a multilevel structure and one or more variables are systematically missing. This systematic missing data pattern may commonly occur in meta-analysis of individual participant data, where some variables are never observed in some studies, but are present in other hierarchical data settings. In these cases, valid imputation must account for both relationships between variables and correlation within studies. Proposed methods for multilevel imputation include specifying a full joint model and multiple imputation with chained equations (MICE). While MICE is attractive for its ease of implementation, there is little existing work describing conditions under which this is a valid alternative to specifying the full joint model. We present results showing that for multilevel normal models, MICE is rarely exactly equivalent to joint model imputation. Through a simulation study and an example using data from a traumatic brain injury study, we found that in spite of theoretical differences, MICE imputations often produce results similar to those obtained using the joint model. We also assess the influence of prior distributions in MICE imputation methods and find that when missingness is high, prior choices in MICE models tend to affect estimation of across-study variability more than compatibility of conditional likelihoods.

Keywords

chained equations; fully conditional specification; joint model; systematic missingness

1. Introduction

Several analytic approaches have been proposed to address missing data with inherent multilevel (hierarchical) structure, but their relative operating characteristics have not, to our knowledge, been extensively studied. Practical guidelines on their use is timely, as an increasing number of analyses are based on data with this structure, where individual observations are nested in larger groups and thus have characteristics that are correlated with members of the same group. These types of data occur in, for example, individual participant data (IPD) meta-analysis of multiple medical studies and joint analysis of individual electronic health records from multiple health systems. Missing data often present

*Correspondence to: kaizar.1@osu.edu.

an unusual challenge in such hierarchical data because some variables may be unavailable in some groups due to group-level design or data availability. This type of “systematic missingness” occurs whenever variables are not measured identically across studies; for example, sites might use different indices to measure socioeconomic status, or record-keeping protocol might differ across hospitals [1]. Because we expect systematic missingness to be prevalent in IPD meta-analysis, we focus on analytical methods for that setting; however, the general principles are more broadly applicable.

When systematic missingness is found in important outcomes, covariates or confounders, appropriate methods are essential to sound analyses. Simply ignoring the missing data, or implementing a complete case analysis, is typically inappropriate, as it could result in biased results even for a missing data mechanism that is missing at random [2]. For systematic missingness, a complete case data analysis is particularly problematic because it would require dropping entire studies from the analysis. If there are enough important systematically missing variables, the complete case approach would result in dropping most or even all datasets from the model, thus undermining the meta-analytic undertaking.

Under the assumption of an ignorable missing data mechanism, multiple imputation is frequently the preferred method of handling missing values since it is often relatively easy to implement, allows use of all available data, and is theoretically unbiased. For the case of multilevel data described above, it is important for the imputation model to correctly reflect the hierarchical covariance structure of the data. For analysis of full IPD data, the meta-analysis literature overwhelmingly promotes the use of mixed-effect models that include a study identifier as a random effect, as this type of model accounts for the variability of individual responses across studies while allowing fixed relationships between covariates in multiple studies (see, e.g., [3]). Recent work has led to similar recommendations for imputation models. For example, Andridge showed that both omitting the effect of study and modeling the effect of study as a fixed effect can lead to bias in estimating regression coefficients even when the intra-class correlation is small [4]; for this reason, random effect models that appropriately account for the hierarchical structure of participants within studies are preferred.

While Andridge’s suggestions are directly applicable to sporadic missingness, systematic missingness complicates imputation models because certain relationships are never observed in some studies. As such, these studies do not contribute to some covariance estimates. Nonetheless, if we assume some across-study homogeneity in the within-study covariance structures, random effect models can be used to impute systematically missing values. In fact, Siddique, et al. demonstrated that IPD meta-analysis with systematic missingness that excludes consideration of hierarchical structure can result in attenuated estimates of regression coefficients [5].

Two types of imputation methods for hierarchical data have been proposed – joint multivariate models for all the variables with missing data [6, 7], and iterative univariate approaches that use multiple imputation with chained equations (MICE), also referred to as “fully conditional specification” (FCS) [1, 8]. Each approach has its advantages and drawbacks. The multivariate modeling approach, although it is a proper multivariate model

for imputation, requires estimating many parameters from the observed data, which can lead to computational limitations and large variances. In addition, specification of the joint model can be quite difficult, especially for extensions that include different types of variables. FCS methods are attractive because they are easy to implement, particularly in cases where a corresponding joint imputation model is not obvious. For example, it is straightforward to impose conditional independence between subsets of variables or, in extensions to the multivariate normal case, specify univariate conditional models for a multivariate response containing many different types of variables. A disadvantage to FCS is that it is often difficult to characterize the theoretical properties of the resulting distribution of the missing values because it makes use of a series of conditional distributions that need not correspond to any joint model. (See, for example, [9].) A notable exception is a simple multivariate normal model, where an appropriately defined FCS has been shown to result in imputation procedures that are equivalent to the joint model [10, 11]. However, this conclusion does not obviously extend to hierarchical normal models. If such a joint model does not exist, it is not clear that the completed data resulting from imputation will have desirable properties. In addition, standard implementations of the univariate approach may not adequately capture correlation induced by the hierarchical data structure.

Multiple imputation implementations specific to random effects models include Schafer [6] and Schafer and Yucel[7]’s joint multivariate normal model, an FCS adaptation of Schafer and Yucel [7]’s approach implemented by van Buuren and Groothuis-Oudshoorn in the “mice” R package [8], and a different FCS implementation proposed by Resche-Rigon, et al[1]. Siddique, et al discussed the use of the joint multivariate model to synthesize a collection of studies with systematic missingness, emphasizing the limitations of this approach when there are many systematically missing variables that are jointly measured in a small number of studies [12]. Resche-Rigon, et al show that their FCS imputation model with random effects outperforms a corresponding fixed effects model when data are systematically missing and normally distributed [1]. Jolani, et al extend the method and results to binary and categorical predictors in a generalized linear model framework and show through simulations that their method has better coverage properties when compared to complete-case analysis, multiple imputation ignoring between-study heterogeneity, and multiple imputation including study as a fixed effect [13]. However, neither Resche-Rigon, et al nor Jolani, et al compared their FCS methods to a joint multivariate model. Enders, Mistler and Keller provide such a comparison for a multivariate normal model with sporadic missing values in the response and conclude based on congeniality arguments that joint model and FCS imputation are both appropriate for random intercepts analysis models, while the two methods are each preferred for contextual effects and random slopes in the analysis model, respectively [14]. van Buuren, et al studied FCS in a multivariate normal model, logistic regression, and quadratic regression [9]. The latter two cases are known to not correspond with a joint model, yet simulation showed that imputations from these models nonetheless achieve nominal confidence levels in all studied cases.

The original formulations for these methods focus on missing individual-level (“level 2”) variables, as opposed to missing variables that are common to all members of a group, or “level 1” variables. We also focus on this situation. However, we note that several authors have studied imputation methods for missing ‘level 1’ variables [15, 16, 17]. In this setting,

Kline, Andridge and Kaizar showed that aggregating “level 2” variables to construct a FCS conditional regression models for each “level 1” variable (as proposed by [15]) results in underestimates of the association between the missing variables and the observed level 2 covariates [17]. This result suggests that for similar multilevel models, it may prove difficult to appropriately capture a complex dataset’s covariance structure using an FCS approach, whereas corresponding joint models can reflect structured relationships among variables.

We fill the gap in direct comparisons of joint and FCS imputation procedures for multilevel models by comparing the joint multivariate model proposed by Schafer [6] with the two implementations of FCS for imputing multilevel data with systematic missingness. We particularly focus on the multivariate normal case because of its analytic tractability. We believe that trends observed in this case are likely to extend to or be exacerbated by other families of distributions. We discuss some cases where an iterative approach will not be exactly equivalent to the full joint model and examine the consequences of these differences in practice. The paper is organized as follows: Section 2 presents the joint model and the corresponding univariate conditional models, and Section 3 discusses the question of exact equivalence between the two models. In Section 4, we present simulation studies to demonstrate the comparative differences between joint and FCS models, and demonstrate practical differences in a meta-analysis of pediatric traumatic brain injury studies in Section 5. Finally, we discuss the implication of our results and make practical recommendations in Section 6. Because our presentation includes notation for the detailed specification of two separate models, we include a notation summary as an Appendix. R code for the simulation and analysis can be found at <https://zenodo.org/badge/latestdoi/92626197>.

2. Imputation model for multilevel missing values

Mimicking a usual IPD meta-analysis setting, we assume the data are consistent with a hierarchical model with two levels – “level-2” individuals nested within “level-1” groups or studies. Specifically, we consider m studies indexed by $i = 1, \dots, m$, each of which contains n_i individuals. Among the collection of studies, there are $r - 2$ systematically missing variables Y that are not observed for any individual in one or more study, so that a joint imputation model for each study i has a matrix response: $\mathbf{Y}_i = (Y_{i1}, \dots, Y_{ir}) \in \mathbb{R}^{n_i \times r}$, where Y_{ir} is a vector of individual measurements of variable r for all n_i individuals in study i . Following [7], we specify a random effects model for $\mathbf{Y}_i$. Conditional on (fully observed) covariates $\mathbf{X}_i \in \mathbb{R}^{n_i \times p}$ and $\mathbf{Z}_i \in \mathbb{R}^{n_i \times q}$, the joint imputation model for an r -dimensional response variable can be written as:

$$\mathbf{Y}_i = \mathbf{X}_i \boldsymbol{\beta} + \mathbf{Z}_i \mathbf{B}_i + \boldsymbol{\varepsilon}_i, \quad (1)$$

where the response $\mathbf{Y}_i$ and error $\boldsymbol{\varepsilon}_i$ are $n_i \times r$ matrices, $\boldsymbol{\beta} \in \mathbb{R}^{p \times r}$ is the matrix of fixed effects, and $\mathbf{B}_i \in \mathbb{R}^{q \times r}$ is the matrix of random effects. The matrices $\mathbf{X}_i$ and $\mathbf{Z}_i$ may contain some or all of the same explanatory variables. Again following typical IPD meta-analytical practice, we assume independent normal distributions on the random effects and errors:

$$\text{vec}(\mathbf{B}_i) \sim N_{qr}(\theta, \Psi); \varepsilon_{ij} \sim N_r(\theta, \Sigma), \quad j=1, \dots, n_i; \quad \mathbf{B}_i \perp \varepsilon_i,$$

where the $\text{vec}(\mathbf{B}_i)$ function stacks the columns of the matrix $\mathbf{B}_i$ on top of each other to create an rq -length column vector. Under this assumption, the distribution of $\mathbf{Y}_i$ (not conditional on the random effects) can be written in vectorized form as a $N_i = m_i$ -dimensional normal distribution:

$$\text{vec}(\mathbf{Y}_i) \sim N(\text{vec}(X_i\beta), \sum \otimes \mathbf{I}_{n_i} + \Psi \otimes Z_i Z_i^T); \quad i=1, \dots, m,$$

where $\mathbf{I}_{n_i}$ is the $n_i \times n_i$ identity matrix and $\otimes$ denotes the Kronecker product.

We follow standard random effects modeling to assume that $\mathbf{B}_i$ and ε_{ij} are identically distributed across studies and $\mathbf{Y}_i$ is conditionally independent of $\mathbf{Y}_{i'}$ for $i \neq i'$. The distribution above fully specifies the joint imputation model for all data with potentially systematically missing values and is parameterized by the matrices β , Σ , and Ψ , which together we denote θ . Under this model, the study-specific variables with systematic missingness, $\mathbf{Y}_i$, each have identical error covariances and the random effect covariances differ across studies only through differences in the random effect covariates Z_i .

To concretely illustrate ideas, we also consider a special case of model (1) where $X_i = Z_i = \mathbf{1}_{n_i}$ a column vector of all ones, and $\mathbf{B}_i$ and β are both row vectors of length r . That is, this random intercepts model assumes a common mean β across all studies with an additional random (mean zero) study effect vector for each study-specific $\mathbf{Y}_i$. When $r=2$, the joint model for $\mathbf{Y}_i$ can be written:

$$\text{vec}(\mathbf{Y}_i) \sim N\left(\beta \otimes \mathbf{1}_{n_i}, \begin{pmatrix} \sigma_{11}\mathbf{I}_{n_i} + \psi_{11}\mathbf{J}_{n_i} & \sigma_{12}\mathbf{I}_{n_i} + \psi_{12}\mathbf{J}_{n_i} \\ \sigma_{12}\mathbf{I}_{n_i} + \psi_{12}\mathbf{J}_{n_i} & \sigma_{22}\mathbf{I}_{n_i} + \psi_{22}\mathbf{J}_{n_i} \end{pmatrix}\right); \quad i=1, \dots, m \quad (2)$$

where $\mathbf{J}_{n_i}$ is a $n_i \times n_i$ matrix of ones.

Model (1) and its special case (2) can be considered to be the true data generation model for $\mathbf{Y}$ conditional on $\mathbf{X}$ and $\mathbf{Z}$. We now turn to imputation models, which may reflect the conditional distribution of the missing values of $\mathbf{Y}$, loosely denoted $\mathbf{Y}^{mis}$ given the observed values of $\mathbf{Y}$, loosely denoted $\mathbf{Y}^{obs}$, and fully observed covariates $\mathbf{X}$ and $\mathbf{Z}$.

2.1. Joint imputation

Following [7], our joint multivariate imputation model begins with the complete data likelihood specified in (1). To create a proper imputation procedure that reflects all the variability in the conditional distribution of $\mathbf{Y}^{mis}$, we additionally specify prior distributions for the parameters in model (1). We denote prior distributions for the covariance matrices Σ and Ψ and the regression coefficients β with the generic function π^J , where the superscript signifies that it is part of the joint imputation model. A Gibbs sampler algorithm can be used to produce imputations from the marginal posterior distribution of the missing data $\mathbf{Y}^{mis}$.

$\mathbf{Y}^{obs}$ by sequentially sampling from the full conditional distributions of the parameters θ , random effects $\mathbf{B}$, and missing data $\mathbf{Y}^{mis}$. One sample can be represented as follows, where the variables in the conditioning set are either observed or the most recently sampled value in the sequential sample:

$$\theta \sim P^\theta(\theta | \mathbf{Y}^{obs}, \mathbf{Y}^{mis}, \mathbf{B}, \mathbf{X}, \mathbf{Z})$$

$$\mathbf{B} \sim P^B(\mathbf{B} | \theta, \mathbf{Y}^{obs}, \mathbf{Y}^{mis}, \mathbf{X}, \mathbf{Z}) \quad (3)$$

$$\mathbf{Y}^{mis} \sim P^Y(\mathbf{Y}^{mis} | \theta, \mathbf{B}, \mathbf{Y}^{obs}, \mathbf{X}, \mathbf{Z})$$

Schafer and Yucel [7] provide a specific sampling algorithm to implement this for the imputation model (1). The specified prior distributions are flat on the mean parameters β and conjugate inverse-Wishart (IW) priors for the variance parameters Σ and Ψ . These prior choices result in full conditional distributions that can each be sampled in closed form. Software that implements this method with Schafer and Yucel's prior distributions is available in the “pan” package in R [18].

2.2. FCS imputation overview

A chained equations or fully conditional specification method is an alternative approach to imputing missing individual values in r variables by specifying a reasonable conditional model (often a generalized linear model) for each variable with missing values given the remaining observed or imputed variables. Each of the r conditional models, denoted g_k , $k = 1, \dots, r$, is parameterized by a variable-specific object θ_k^* (analogous to θ in the joint imputation model; we use asterisks to differentiate entities that appear in the FCS models from entities that have similar interpretations in the joint model). Upon choosing prior distributions π_k^{FCS} for each θ_k^* , one sample from the FCS imputation sequential sampling algorithm for a random effects model is drawn as follows (loosely following notation found in [11] and [19]):

$$\begin{aligned} \theta_1^*, \mathbf{B}_{(1)}^* &\sim \tilde{g}_1(\theta_1^*, \mathbf{B}_{(1)}^* | \mathbf{Y}, \mathbf{X}, \mathbf{Z}) \\ \mathbf{Y}_{(1)}^{mis} &\sim g_1(\mathbf{Y}_{(1)}^{mis} | \mathbf{Y}_{(1)}^{obs}, \mathbf{Y}_{(-1)}, \theta_1^*, \mathbf{B}_{(1)}^*, \mathbf{X}, \mathbf{Z}) \\ &\dots \\ \theta_r^*, \mathbf{B}_{(r)}^* &\sim \tilde{g}_r(\theta_r^*, \mathbf{B}_{(r)}^* | \mathbf{Y}, \mathbf{X}, \mathbf{Z}) \\ \mathbf{Y}_{(r)}^{mis} &\sim g_r(\mathbf{Y}_{(r)}^{mis} | \mathbf{Y}_{(r)}^{obs}, \mathbf{Y}_{(-r)}, \theta_r^*, \mathbf{B}_{(r)}^*, \mathbf{X}, \mathbf{Z}) \end{aligned} \quad (4)$$

where $\mathbf{Y}_{(-k)}$ denotes $\mathbf{Y}$ with the k th variable removed, $\mathbf{Y}_{(k)}$ denotes the portion of $\mathbf{Y}$ that includes only the k th variable, and $\mathbf{B}_{(k)}^*$ is a $m \times q^*$ random effect matrix associated with variable k .

The full conditional distribution for the parameters $\tilde{g}_k$ is determined by the prior π_k^{FCS} and full conditional likelihood g_k . When proper priors are used, sequential samples from each pair of full conditional distributions $\{\tilde{g}_k, g_k\}$ would produce draws from a well-defined predictive distribution of $\mathbf{Y}_{(k)}^{\text{mis}}$ conditional on fixed $\mathbf{Y}_{(k)}^{\text{obs}}, \mathbf{Y}_{(-k)}$. As noted in the introduction, however, neither the functional form of g_k nor its corresponding parameterization needs to correspond to a joint r -variate distribution on the full array $\mathbf{Y}$. As a consequence, there is no guarantee that the variables in $\mathbf{Y}^{\text{mis}}$ are drawn from any proper joint imputation distribution. This is a well-known potential limitation of such algorithms, and their behavior in this area is understood in relatively few particular cases.

2.3. FCS implementation

The FCS algorithms we consider use a series of univariate mixed-effect regression models to impute missing values one variable at a time. With model (1) in mind, the conditional distribution g_k for the k th variable is the product of study-specific conditional distributions, g_{ik} . In turn, these are specified as random effects regression models with conditional mean $X_{ik}^* \beta_k^*$ and a study-specific random effect $B_{(k)i}^*$:

$$g_k(\mathbf{Y}_{(k)}^{\text{mis}} | \mathbf{Y}_{(k)}^{\text{obs}}, \mathbf{Y}_{(-k)}, \mathbf{X}, \mathbf{Z}, \mathbf{B}_{(k)}^*, \theta_k^*) = \prod_{i \in \mathcal{M}_k} g_{ik}(Y_{ik} | \mathbf{X}_{ik}^*, \mathbf{Z}_{ik}^*, B_{(k)i}^*, \theta_k^*) \quad (5)$$

$$g_{ik}(Y_{ik} | \mathbf{X}_{ik}^*, \mathbf{Z}_{ik}^*, B_{(k)i}^*, \theta_k^*) = \varphi(\mathbf{X}_{ik}^* \beta_k^* + \mathbf{Z}_{ik}^* B_{(k)i}^*, \sigma_k^* I), k=1, \dots, r$$

where $\mathcal{M}_k$ is the collection of indices for studies where variable k is systematically missing, Y_{ik} is a length n_i vector, $\mathbf{X}_{ik}^*$ is a $n_i \times p^*$ matrix that may include $\mathbf{Y}_{i(-k)}$ or transformations of $\mathbf{Y}_{i(-k)}$ and β_k^* is a length p^* vector, $\mathbf{Z}_{ik}^*$ is a $n_i \times q^*$ matrix that also may include $\mathbf{Y}_{i(-k)}$ or transformations of $\mathbf{Y}_{i(-k)}$ and $B_{(k)i}^*$ is a length q^* vector, and $\varphi(\cdot, \cdot)$ is the multivariate normal density function. The random effects $B_{(k)i}^*$ are independently and identically distributed (iid) as $N(0, \psi_k^*)$. Each g_k is parameterized by the collection $\theta_k^* = \{\beta_k^*, \sigma_k^*, \psi_k^*\}$.

With the special case model (2) in mind, the fully conditional models g_{i1} and g_{i2} for $\mathbf{Y}$ consistent with (5) can be written as regressions with random intercepts, where for this special case we follow usual practice to define the matrices $\mathbf{X}_{i1}^* = (\mathbf{X}_i, Y_{i2})$ and $\mathbf{X}_{i2}^* = (\mathbf{X}_i, Y_{i1})$.

$$g_{i1}(Y_{i1} | Y_{i2}, X_i, B_{(1)i}^*, \theta_1^*) = \varphi(\beta_1^* \mathbf{X}_{i1}^* + B_{(1)i}^* \mathbf{1}_{n_i}, \sigma_1^* \mathbf{I}_{n_i}); B_{(1)i}^* \stackrel{\text{iid}}{\sim} N(0, \psi_1^*) \quad (6)$$

$$g_{i2}(Y_{i2} | Y_{i1}, X_i, B_{(2)i}^*, \theta_2^*) = \varphi(\beta_2^* \mathbf{X}_{i2}^* + B_{(2)i}^* \mathbf{1}_{n_i}, \sigma_2^* \mathbf{I}_{n_i}); B_{(2)i}^* \stackrel{\text{iid}}{\sim} N(0, \psi_2^*)$$

Each multilevel conditional model implies a structured variance-covariance matrix when the latent random effects are removed from the conditioning set: the covariance of $Y_{11}|Y_{12}$, for instance, is equal to $\sigma_1^* \mathbf{I}_{n_i} + \psi_1^* \mathbf{J}_{n_i}$.

We consider two implementations of FCS using the imputation model described in (4) and (5) above. These two use the same likelihood functions g_k but differ in the procedure used to sample θ_k^* from $\tilde{g}_k$ in the first step of the algorithm in (4) above. The first method is presented by Resche-Rigon, et al and at each iteration samples θ_k^* from the estimated asymptotic sampling distribution of its maximum likelihood estimator [1]. The second method is the MICE implementation of multilevel normal imputation as available in the R “mice” package [8] using the “2l.impute.pan” imputation function. The next sections provide additional detail about these two methods, and in Section 3.3, we compare the implications of the differences in prior distributions. To differentiate among the general and specific features of FCS models and implementations, when our discussion applies to any MICE algorithm using the model (6), we refer to the method as simply MICE or FCS. When the distinction between these two implementations is important, we will refer to the method in [1] as “MICE-ML” and the “2l.impute.pan” method in the “mice” R package as “MICE-IW”.

2.3.1. MICE-ML FCS implementation—For each sequential sample in the MICE-ML method, one first calculates the maximum likelihood estimator $\hat{\theta}_k^*$ of θ_k^* conditional on $\mathbf{X}_k^*$ and $\mathbf{Z}_k^*$ and imputes each variable using the full conditional likelihood g_k in (5). In the $r=2$ case, the steps for drawing $(\theta_1^*, B_{(1)l})$ from $\tilde{g}_1$ corresponding to the likelihood g_1 shown in (6) above are:

$$\beta_1^* \sim N(\hat{\beta}_1^*, \widehat{\text{Var}}(\hat{\beta}_1^*))$$

$$\begin{pmatrix} \sigma_1^* \\ \psi_1^* \end{pmatrix} \sim N\left(\begin{pmatrix} \hat{\sigma}_1^* \\ \hat{\psi}_1^* \end{pmatrix}, \widehat{\text{Var}}\left(\begin{pmatrix} \hat{\sigma}_1^* \\ \hat{\psi}_1^* \end{pmatrix}\right)\right)$$

$$B_{(1)i} \sim N(0, \psi_1^*); \quad i=1, \dots, m$$

where $\widehat{\text{Var}}\left(\begin{pmatrix} \hat{\sigma}_1^* \\ \hat{\psi}_1^* \end{pmatrix}\right)$, the estimated covariance matrix of $(\hat{\sigma}_k^*, \hat{\psi}_k^*)$, is obtained by evaluating the Fisher information at $(\hat{\sigma}_k^*, \hat{\psi}_k^*)$. This method specifies no prior information about θ_k^* , instead using the data alone to define $\tilde{g}_k$. Once such an algorithm converges, the steps here will approximate a draw from the posterior distribution of $(B_{(1)l}, \theta_1^*)$ given $(\mathbf{Y}_{(1)}^{\text{obs}}, \mathbf{Y}_{(2)}, \mathbf{X}, \mathbf{Z})$. Sampling from these data-driven full conditional distributions raises two practical problems. First, the estimated variance-covariance matrix $\widehat{\text{Var}}\left(\begin{pmatrix} \hat{\sigma}_1^* \\ \hat{\psi}_1^* \end{pmatrix}\right)$ may

not be positive definite, and thus may not be used to generate samples. In this case, Resche-Rigon, et al recommend using the maximum likelihood estimates directly instead of sampling from the normal distribution [1]. Second, draws of σ_1^* and ψ_1^* may not be positive. To prevent this situation, we implement an algorithm that truncates the normal distribution at zero.

2.3.2. MICE-IW Implementation—The MICE-IW method uses existing functionality from the R package “pan” [18], typically used for joint model imputation of multivariate normal models as in section 2.1, to perform univariate imputations. That is, each pair of full conditional univariate distributions $\{\tilde{g}_k, g_k\}$ is defined using the same family of priors and conditional sampling methods as the distributions $\{P^\theta, P^B, P^Y\}$ defined for the joint procedure (3). Under this method, the prior distributions π_k^{FCS} are flat on β_k^* and inverse-Wishart on σ_k^* and ψ_k^* with degrees of freedom and prior scales ν_k and V_k for each variable k . This prior specification results in a closed form expression for the full conditional distributions. For our specific case where $r = 2$, draws of the parameters and random effects $(\theta_1^*, B_{(1)i})$ are sampled as:

$$B_{(1)i} \sim N(\mu(B_{(1)i}), \sigma(B_{(1)i})); \quad \mu(B_{(1)i}) = \frac{\sigma(B_{(1)i})}{\sigma_1^*} (Y_{i1} - \mathbf{X}_{i2}^* \beta_1^*); \quad \sigma(B_{(1)i}) = \left(\frac{1}{\psi_1^*} + \frac{n_i}{\sigma_1^*} \right)^{-1}$$

$$\beta_1^* \sim N \left(\sum_{i=1}^m (\mathbf{X}_{i1}^{*T} \mathbf{X}_{i1}^*)^{-1} \sum_{i=1}^m (\mathbf{X}_{i1}^{*T} (Y_{i1} - 1_2 B_{(1)i})) , \sigma_1^* \sum_{i=1}^m (\mathbf{X}_{i1}^{*T} \mathbf{X}_{i1}^*)^{-1} \right)$$

$$\sigma_1^{*-1} \sim \text{IW} \left(\nu_1 + \sum_{i=1}^m n_i, \left(V_1^{-1} + \sum_{i=1}^m (Y_{i1} - B_{(1)i} 1_{n_i} - Y_{i2} \beta_1^*)^2 \right)^{-1} \right)$$

$$\psi_1^{*-1} \sim \text{IW} \left(\nu_2 + m, \left(V_2^{-1} + \sum_{i=1}^m B_{(1)i}^2 \right)^{-1} \right)$$

These conditional draws will, upon convergence, represent a sample from the posterior distribution $\tilde{g}$ of $(\theta_1^*, B_{(1)i})$ as determined by the likelihood g_1 and prior distributions π_1^{FCS} . The algorithm for drawing $(\theta_2^*, B_{(2)i})$ is similar.

3. Equivalence of joint and conditional imputation models

If the joint and FCS imputation procedures can be shown to be equivalent for the normal hierarchical model, then one could choose the most convenient algorithm without being concerned about relative statistical properties of the procedures. If not, an understanding of situations where differences are substantial will be helpful in guiding algorithmic choices.

Imputation procedures are equivalent when the posterior distribution of the missing variables used to sample imputed values, or imputation distribution, is identical across procedures. Studying such equivalence, or lack of equivalence, can be segmented into three considerations. First, because FCS procedures are ad hoc specifications, any possible equivalence relies on the existence of a stationary posterior distribution in the FCS algorithm. Second, if such a posterior distribution exists, it must be compatible with the joint imputation model in order for the two imputation methods to be asymptotically equivalent [11]. Finally, the choice of prior distributions may impact finite sample imputations.

We describe these three items in more detail below, and address each for the joint and FCS procedures for imputing variables in the context of the hierarchical multivariate normal data model. In particular, we compare the two imputation methods given by a fully Bayesian joint model, described in 3), and the analogous MICE procedure (4). For ease of presentation, we focus on the special two-dimensional case shown in (2) and (6) above. As noted in the text below, we sometimes also assume that $\beta = 0$; i.e., we assume there are no fixed effects in the joint model and that $E(Y_{ik}) = 0$ for $k = 1, 2$. Throughout, we continue to denote the data distribution (i.e., the likelihood) for the joint model with $P(Y|\theta)$ where $\theta = (\beta, \Sigma, \Psi)$. The MICE conditional models for the data are denoted g_1, g_2 and are parameterized by $\theta_1^* = \{\beta_1^*, \sigma_1^*, \psi_1^*\}$ and $\theta_2^* = \{\beta_2^*, \sigma_2^*, \psi_2^*\}$, respectively. The prior distributions are $\pi^J(\theta)$ for the joint model and $\pi_k^{\text{FCS}}(\theta_k^*)$; $k = 1, 2$ for MICE. We assume that the model for the data $P(Y|\theta)$ implied by (2) is correctly specified, so that imputations from the joint model will lead to consistent estimates in the final model.

3.1. Existence

The sequence of conditional draws in an FCS method must converge to some stationary posterior distribution in order for it to be meaningful to evaluate its implied imputation distribution. Arnold and Press have shown that two conditional densities g_1 and g_2 on y_1 and y_2 , respectively, will together specify a joint distribution if and only if their ratio $\frac{g_1(y_1|y_2)}{g_2(y_2|y_1)}$ can be factorized as $g(y_1)h(y_2)$ [10]. (See also [20].) For FCS conditional models (6), this factorization condition will hold when the conditional parameters θ_k^* are functions of θ as necessary for compatibility with P , as discussed in the next subsection. In this case, the unique joint distribution will be the distribution given by (2).

3.2. Compatibility of models (2) and (6)

Loosely speaking, a full conditional imputation approach would be called compatible with a joint model if the full conditional distributions for the missing data are identical to those that would be derived from the joint model, and their parameters can be derived from the parameters for the joint model. To more precisely define compatibility, we follow definitions given in [11] and [21] and first define $P_k(Y_{(k)}|Y_{(-k)}, \theta_k)$ to be the full conditional distribution of the variable $Y_{(k)}$ that is derived from the full joint model $P(Y|\theta)$. This conditional distribution is parameterized by θ_k , possibly a subset of θ . Then, a family of conditional models $\{g_1, \dots, g_r\}$, where g_k is parameterized by θ_k^* , is said to be compatible with P if $g_k = P_k$ for all k , and θ is surjective with $\theta^* = (\theta_1^*, \dots, \theta_r^*)$ [11].

Compatibility is an important consideration for evaluating imputation procedures because if the FCS distributions $\{g_1, \dots, g_r\}$ are not compatible with any joint model, the posterior predictive distribution of the resulting imputations will generally not correspond to any Bayesian model. Compatibility implies the asymptotic equivalency of imputation distributions for Y^{mis} , whenever suitable asymptotic conditions ensure the choice of prior distributions for the model parameters θ have no asymptotic impact [11].

We first consider the compatibility requirement that $g_k = P_k$ for all k when P is the multivariate normal model in (2) and the FCS analogs $\{g_1, g_2\}$ are given in (6). Let P_1, P_2 denote the conditional distributions of Y_{i1} and Y_{i2} based on P . The following expressions give the conditional mean and covariance of $P_1(Y_{i1}|Y_{i2}, \theta_1)$ when the mean of the joint distribution is zero, i.e., $\beta = 0$:

$$E(Y_{i1}|Y_{i2}, \theta_1) = (\beta_{1.2,1} \mathbf{I}_{n_i} + \beta_{1.2,2}^i \mathbf{J}_{n_i}) Y_{i2} \quad (7)$$

$$\text{Cov}(Y_{i1}|Y_{i2}, \theta_1) = \sigma_{1.2} \mathbf{I}_{n_i} + \psi_{1.2}^i \mathbf{J}_{n_i}$$

where the parameters $(\beta_{1.2,1}, \beta_{1.2,2}^i, \sigma_{1.2}, \psi_{1.2}^i)$ are functions of the covariance matrix of Y_i , as defined below. We use the subscript notation $k.k'$ to denote a parameter corresponding to the conditional distribution $P_k(Y_{(k)}|Y_{(k')}, \theta)$, and as necessary, additional indices in the subscript identify subsets of a multidimensional parameter. The correspondence of these conditional parameters to those of the joint distribution are as follows:

$$\beta_{1.2,1} = \frac{\sigma_{12}}{\sigma_{22}}$$

$$\beta_{1.2,2}^i = \frac{\sigma_{22}\psi_{12} - \sigma_{12}\psi_{22}}{\sigma_{22}(\sigma_{22} + n_i\psi_{22})} \quad (8)$$

$$\sigma_{1.2} = \sigma_{11} - \frac{\sigma_{12}^2}{\sigma_{22}}$$

$$\psi_{1.2}^i = \psi_{11} + \frac{n_i\sigma_{22}\psi_{12}^2 + 2\sigma_{12}\sigma_{22}\psi_{12} - \sigma_{12}^2\psi_{22}}{\sigma_{22}(\sigma_{22} + n_i\psi_{22})}$$

The expressions in (7) indicate that, as long as the FCS parameter $\theta_1^* = \{\beta_1^*, \sigma_1^*, \psi_1^*\}$ is equal to the joint parameter $\theta_1 = (\beta_{1.2}^i, \sigma_{1.2}, \psi_{1.2}^i)$ defined in (8), it is possible to express the full conditional distribution P_k as a mixed-effects regression model of the form of the FCS

model (6). The multilevel structure of this model, however, creates two unusual features in the conditional distributions. First, the conditional mean is not a simple linear regression model. The $\beta_{1.2,2}^i \mathbf{J}_{n_i}$ term creates a multilevel structure in the conditional mean, so that for individual j , the mean of $Y_{1j} | Y_{2j}$ is linear in both the j th value of Y_{2j} and the within-study sum $\sum_{j=1}^{n_i} Y_{i2j}$, or, equivalently, its within-study mean. An FCS model g_k must include this additional within-study sum as a covariate in the regression function in order to be compatible with model (2). An analogous result holds for models with a random slope associated with a continuous predictor Z_i : the conditional mean is linear in $Z_i' Y_{i2}$ and compatibility requires that this term should be included in $\mathbf{X}_{i1}^*$ as a fixed effect.

Carpenter and Kenward note the need to include the study-specific mean, and show that this implies exact equality of P_k and g_k in the case of equal numbers of “level 1” measurements per “level 2” group. (i.e., $n_j = n$ for all j) [22]. While this is a reasonable assumption for the analysis of a single repeated measures study, it becomes untenable for IPD meta-analysis or hierarchical models that include random effects. When study sizes vary, the conditional distributions are heterogeneous across studies: Equations (8) show that although the joint model specifies homogeneous error and random effect variances, its corresponding conditional distribution P_1 has parameters that depend on $Z_i' Z_i$, which need not be the same for each study. In particular, for this random intercept model, the parameters of $\beta_{1.2,2}^i$ and $\psi_{1.2}^i$ depend on $\mathbf{1}_{n_i}^T \mathbf{1}_{n_i} = n_i$, and a model assuming these values are constant across study cannot be compatible with the model (2).

Figure 1 depicts the practical influence of heterogeneous sample size on the magnitude of these two parameters, $\beta_{1.2,2}^i$ (left panel) and $\psi_{1.2}^i$ (right panel) for a case where the joint model covariance matrices Σ and Ψ are of similar magnitude to those estimated from the TBI data in Section 5. In each plot, the horizontal axis represents the study size and the vertical represents the value of the conditional parameter for a study of that value of n_j . Each line shows the effect of increasing sample size for one particular value of the intraclass

correlation in Ψ , defined as $\rho = \frac{\psi_{12}}{\sqrt{\psi_{11}\psi_{22}}}$. The dashed line represents a case where Σ and Ψ have similar correlations (0.65 and 0.7, respectively), and both plots show nearly flat lines, indicating that for this case the parameters vary little with study size. Conversely, when the random effects are extremely negatively correlated ($\rho = -0.90$), the conditional distributions can differ dramatically across studies with different sample sizes. Consequently the MICE model (6) that assumes complete homogeneity of conditional distributions is theoretically incompatible with P . The magnitude of differences across conditional distributions depends on both the model covariances and the variability of n_j across studies (or, if a random slope is used, the across-study variability of $Z_i' Z_i$). For both parameters, the lines tend to grow flat for n_j sufficiently large; the effect of sample size on the parameters diminishes as studies grow.

When some variables are systematically missing, it is nonetheless necessary to assume some homogeneity in the coefficients across studies for FCS formulations that are based on

standard generalized linear models. In a practical sense, then, compatibility will not be possible. The simulation study in Section 4 examines the potential consequences of this trade-off in several situations.

In addition to equivalent forms of P_k and g_k , compatibility requires a surjective map t that maps the joint parameterization to the conditional parameterization. That is, if g_k is compatible with P , it must be parameterized by a valid value of $\kappa(\theta)$. The reparameterization t for the simple two-variable case with $\beta = 0$ is given in (8). The inverse of t maps an 8-dimensional parameter $\theta^* = \{\theta_1^*, \theta_2^*\}$ to two covariance matrices $\theta = \{\Sigma, \Psi\}$. Because these are positive definite matrices, θ has only 6 free parameters, one of which must satisfy an additional constraint that ensures a positive determinant of Ψ . As a result, each θ_k^* lies in a specific subspace of R^4 , and values outside of this space will not correspond to a valid value of θ . In practice, draws of these parameters should fall in the correct space with high probability for large sample sizes. It is nonetheless possible for an FCS algorithm to draw parameter values outside of the correct subspace, particularly if the chosen priors put high probability on incompatible θ^* values.

3.3. Priors

If the FCS models $\{g_1, \dots, g_r\}$ are compatible with a joint model P , both methods will draw imputations from a posterior predictive distribution of $Y^{mis} | Y^{obs}$. For finite samples, however, each method's predictive distribution will depend on the prior distributions placed on the model parameters, $\pi^{FCS}(\theta^*)$ and $\pi^J(\theta)$. Liu, et al show that under several technical conditions, these differences will disappear asymptotically [11]. In many multilevel applications, however, asymptotic results are unlikely to be relevant. For example, in an IPD meta-analysis synthesizing a small number of studies, the prior distribution on the model's study-level parameters may be particularly influential, and the asymptotic result would depend on unrealistic increases in the number of studies. A practical comparison of the methods, then, requires an assessment of how the priors influence imputations in finite samples.

To aid in comparison of π^{FCS} and π^J , we first reconcile the two parameterizations of the model. Note that when the models are compatible, for any k , the joint distribution P can be written $P(Y|\theta) = g_k(Y_{(k)} | Y_{(-k)}, \theta_k^*) P(Y_{(-k)} | \tilde{\theta}_k)$, where g_k is the compatible FCS distribution parameterized by θ_k^* , and $(\theta_k^*, \tilde{\theta}_k)$ is a function of θ [11]. It is possible to transform any prior distribution on the joint parameterization $\pi^J(\theta)$ to find the corresponding prior $\pi^J(\theta_k^*, \tilde{\theta}_k)$. Integrating $\tilde{\theta}_k$ out of π^J gives the joint model prior distribution on θ_k^* that can be readily compared to the FCS prior on the same parameter. We make this comparison using the θ_k^* parameterization because this involves only specifying the marginal priors $\{\pi^{FCS}(\theta_k^*)\}_{k=1}^r$, while the reverse comparison would require specifying a joint prior on $(\theta_1^*, \dots, \theta_r^*)$.

We assume the FCS model (4) is chosen to be compatible with (2), so that the fixed-effect design X_{ik}^* includes $(Y_{i(-k)}, 1_{n_i} \sum_{j=1}^{n_i} Y_{i(-k)j})$ as discussed in Section 3.2. Then the joint and

conditional parameters can be represented, following the notational conventions in (8) and again fixing $\beta = 0$:

$$\theta = (\Sigma, \Psi) \quad (9)$$

$$\theta_1^* = (\beta_{1.2,1}, \beta_{1.2,2}, \sigma_{1.2}, \psi_{1.2}); \quad \tilde{\theta}_1 = (\sigma_{22}, \psi_{22}) \quad (10)$$

$$\theta_2^* = (\beta_{2.1,1}, \beta_{2.1,2}, \sigma_{2.1}, \psi_{2.1}); \quad \tilde{\theta}_2 = (\sigma_{11}, \psi_{11}) \quad (11)$$

It is straightforward to show the one-to-one correspondence between θ and $(\theta_k^*, \tilde{\theta}_k)$. We consider the prior distributions for the parameters that were proposed for the three imputation procedures described in Section 2.3 above. First, we look at weakly-informative inverse-Wishart prior on θ as proposed by [6] and implemented in the ‘pan’ R package [18]. We compare this prior on θ to two sets of FCS priors on θ_k^* : a univariate analog to the inverse-Wishart priors on θ used in the MICE-IW method, and the implicitly non-informative prior that leads to the large-sample conditional sampling distributions used in the MICE-ML method. Our comparison of these priors considers only θ_2^* , omitting the redundant study of θ_1^* . We begin by precisely describing the three sets of prior distributions, then examine the shape and influence of the various prior distribution specifications.

Joint Model Prior Distributions—In their fully Bayesian proposal for imputation within a multilevel model, Schafer and Yucel use conditionally conjugate and *a priori* independent inverse-Wishart prior distributions on the covariance matrices Σ and Ψ and improper flat prior distributions on fixed regression coefficients β [7]. An inverse-Wishart (IW($\nu, \mathbf{V}$)) prior has two hyperparameters: the scale matrix $\mathbf{V}$ and the degrees of freedom ν . For our two-variable example, we use the “minimally informative” $\nu = 2$, and choose prior scale as $\mathbf{V} = \mathbf{I}_2$ to follow the default implementation used in the analogous MICE-IW procedure. The joint prior distribution for $\theta = (\Sigma, \Psi)$ under these settings, up to a normalizing constant, is:

$$\pi^J(\theta) \propto |\Sigma|^{-5/2} |\Psi|^{-5/2} \exp \left(-\frac{1}{2} \left(\frac{\sigma_{11} + \sigma_{22}}{|\Sigma|} + \frac{\psi_{11} + \psi_{22}}{|\Psi|} \right) \right) \quad (12)$$

This distribution implies a prior distribution in the FCS parameterization:

$$\begin{aligned}
\pi^J(\theta_2^*, \tilde{\theta}_2) &\propto |\sigma_{11}(n\psi_{11} + \sigma_{11})|(\sigma_{11}\sigma_{2.1})^{-5/2}(-\beta_{2.1,2}^2\sigma_{11}(n\psi_{11} + \sigma_{11}) + \psi_{2.1}\psi_{11})^{-5/2} \\
&\exp\left(\frac{1}{2}\frac{(\beta_{2.1,2}\psi_{11}n^2 + 2n\beta_{2.1,2}\beta_{2.1,1}\psi_{11} + n\beta_{2.1,2}^2\sigma_{11} + \beta_{2.1,1}\psi_{11} + \psi_{2.1} + 2\beta_{2.1,2}\beta_{2.1,1}\sigma_{11} + \psi_{11})}{\beta_{2.1,2}^2\sigma_{11}(n\psi_{11} + \sigma_{11}) - \psi_{2.1}\psi_{11}}\right) \\
&\exp\left(-\frac{1}{2}\left(\frac{1}{\sigma_{2.1}}\left(1 + \frac{\sigma_{11}}{\sigma_{2.1}\beta_{2.1,1}^2\sigma_{11}}\right)\right)\right) \mathcal{I}\left(|\beta_{2.1,2}| < \sqrt{\frac{\psi_{2.1}\psi_{11}}{\sigma_{11}(n\psi_{11} + \sigma_{11})}}\right)
\end{aligned}
\tag{13}$$

where $\mathcal{I}$ is the indicator function.

MICE-IW Prior Distributions—The R “mice” package implements an analogous FCS algorithm that uses univariate versions of the inverse-Wishart joint model prior distributions for the parameters in each full conditional specification to iteratively generate imputations one variable at a time. That is, each step of the FCS algorithm implements the joint imputation model proposed by [6] as if only one variable in the dataset had missing values. As such, the prior distribution $\pi^{\text{FCS}}(\theta_2^*)$ has the same functional form as π^J , but differs in dimension and the parameterization to which it applies. MICE-IW imposes independent improper flat distributions on each column of $\beta_{2.1}$, and independent inverse Chi-squared distributions with $\nu = 1$ degree of freedom for the variance parameters $\psi_{2.1}$ and $\sigma_{2.1}$. This prior distribution for the vector θ_2^* can be written:

$$\pi^{\text{FCS}}(\theta_2^*) \propto \sigma_{2.1}^{-3/2} \psi_{2.1}^{-3/2} \exp\left(-\frac{1}{2}(\psi_{2.1}^{-1} + \sigma_{2.1}^{-1})\right) I(\psi_{2.1} > 0; \sigma_{2.1} > 0) \tag{14}$$

MICE-ML Prior Distributions—The MICE method in [1] does not explicitly specify priors for θ_2^* , but instead directly specifies their full conditional distributions for the sequential sampling procedure:

$$\begin{pmatrix} \beta_{2.1,1} \\ \beta_{2.1,2} \end{pmatrix} \sim N\left(\begin{pmatrix} \hat{\beta}_{2.1,1} \\ \hat{\beta}_{2.1,2} \end{pmatrix}, \widehat{\text{Var}}\begin{pmatrix} \hat{\beta}_{1.2} \\ \hat{\beta}_{2.2} \end{pmatrix}\right) \tag{15}$$

$$\begin{pmatrix} \sigma_{2.1} \\ \psi_{2.1} \end{pmatrix} \sim N\left(\begin{pmatrix} \hat{\sigma}_{2.1} \\ \hat{\psi}_{2.1} \end{pmatrix}, \widehat{\text{Var}}\begin{pmatrix} \sigma_{2.1} \\ \psi_{2.1} \end{pmatrix}\right), \tag{16}$$

where the estimated parameters $\hat{\beta}_{2.1,1}$, $\hat{\beta}_{2.1,2}$, $\hat{\sigma}_{2.1}$ and $\hat{\psi}_{2.1}$ are maximum likelihood estimates and the estimated variances are asymptotic variance estimates derived from the Fisher information. All are available in closed form for the normal model [23]. This

specification for $\beta_{2,1}$ is equivalent to using an improper flat prior distribution, but it is less straightforward to identify a corresponding prior distribution for the variance parameters.

The absence of a prior distribution here means that it is not necessary to provide subjective information about model parameters, but posterior draws of model parameters are dependent on the maximum likelihood estimates alone. One possible disadvantage to the procedure described above is that it relies on a large-sample distribution of both $\sigma_{1,2}$ and $\psi_{1,2}$, which may not be a close approximation if either the number of studies and number of observations per study are small. In addition, any bias in the maximum likelihood estimator may be present in posterior draws of θ_k^* . But there may be a unique advantage to a data-driven approach: as discussed below, it is possible for standard priors on θ_k^* to put substantial mass in regions where these parameters are incompatible with the joint model. Estimates from data should, with high probability, avoid these regions.

Comparisons among Prior Distributions—Direct comparison of the prior distributions in (13) and (14) would require integrating over σ_{11} and ψ_{11} ; instead, we examine the two expressions and note a few differences. Our inability to factor out the marginal parameters σ_{11} , ψ_{11} suggests that the joint model imposes a relationship between the conditional and marginal parameters that any FCS prior on θ_2^* alone would be unable to model directly. A notable feature of this relationship is the indicator function in (13), which

truncates the support of the parameter $\beta_{2,1,2}$ at $\pm \sqrt{\frac{\psi_{2,1}\psi_{11}}{\sigma_{11}(n\psi_{11}+\sigma_{11})}}$. This restriction is necessary in order for Ψ to be positive definite and often requires that $\beta_{2,1,2}$ be small in magnitude. Note that a similar constraint does not hold for $\beta_{2,1,1}$; $\sigma_{2,1} > 0$ is sufficient to ensure Σ is positive definite. The FCS flat prior samples $\beta_{2,1,2}$ from an unrestricted normal distribution; as a consequence, if this parameter is poorly estimated, sampled values might be outside of the support under π^J . As discussed in more detail in Section 4, simulations suggest that this might lead to unstable estimates when data are limited and the number of studies is small.

Figure 2 shows marginal prior density estimates of the each element of θ_2^* under π^J and, when applicable, the MICE-IW FCS prior distribution for the same parameter. The prior density on $\beta_{2,1,2}$, when averaged over the other parameters, is highly concentrated near zero. The prior for the slope $\beta_{2,1,1}$ is more diffuse than $\beta_{2,1,2}$, but still not uninformative. For both of the variance parameters $\sigma_{2,1}$ and $\psi_{2,1}$, the densities are similarly shaped, with heavier tails in the FCS priors.

To explore the practical effects of the prior differences we simulated values from each posterior predictive distribution of a new observation $\check{Y}_{i1j}$ given one simulated data set Y_1 , Y_2 , with no missing values. These samples are obtained by sampling many values of the parameters θ and θ_1^* from their posterior distributions P^θ and $\tilde{g}_1$, respectively and evaluating the conditional likelihoods P_1 and g_1 at each sampled value. Draws from these likelihoods are samples from the posterior predictive distribution. The quantile-quantile plots in Figure 3 compare the predictive distributions associated with the joint and MICE-IW priors. Each column of the figure shows samples of the predictive distribution of a particular observation

$\check{Y}_{1j}$ conditional on Y_2 . The rows represent varying values of the study size (n) and number of studies (m). These plots show that for small values of m in particular, the joint predictive distributions can often have much heavier tails than the FCS counterpart. The plots in the upper rows show that as samples grow large, these differences are less pronounced. We note that although each plot represents the predictive distribution for only four randomly chosen individuals in the dataset, the results were similar for those not displayed.

4. Simulations

We present five simulation studies that compare joint model imputation to the FCS implementations discussed above. The goal is to assess the impact of incompatibility and prior influence on the performance of these methods when using default software settings. Each study compares four imputation models: the joint model (2) with fully observed predictors and random effects X_i and z_i , MICE-ML as described in section 2.3.1, and MICE-IW as described in section 2.3.2 using two separate definitions of X_{ik}^* , the fixed effects in the conditional regressions. The first uses $X_{ik}^* = (X_i, Y_{i(-k)})$ and the second includes the within-study sums as predictors, so that $X_{ik}^* = (X_i, Y_{i(-k)}, Z_i' Y_{i(-k)})$. The second model ensures theoretical compatibility when each study has equal sample sizes, so we refer to this as “compatible MICE-IW” (or MICE-IW-c) and to the first method as “MICE-IW”. The “MICE-ML” method was also implemented using the compatible regression with $X_{ik}^* = (X_i, Y_{i(-k)}, Z_i' Y_{i(-k)})$. We used R to implement all methods: we used the R “pan” package [18] for the joint imputation, the R “mice” package [8] for MICE-IW, and the “lme4” package to obtain maximum likelihood estimates for the MICE-ML method.

In the first four studies, we simulated datasets with $r = 3$ mean-zero variables as $Y_{ij} = e_{ij} + z_i b_i$, where $e_{ij} \sim N_3((0 \ 0 \ 0)^T, \Sigma)$ and $b_i \sim N_3((0 \ 0 \ 0)^T, \Psi)$. We used twelve combinations of the covariance matrices Σ and Ψ , the first nine of which have the form

$$\Sigma = \sigma \begin{pmatrix} 1 & \rho_1 & 0.6 \\ \rho_1 & 1 & \rho_3 \\ 0.6 & \rho_3 & 1.5 \end{pmatrix}; \quad \Psi = \psi \begin{pmatrix} 1 & \rho_2 & 0.6 \\ \rho_2 & 1 & \rho_4 \\ 0.6 & \rho_4 & 1.5 \end{pmatrix},$$

with $\sigma = 10$ and $\psi = 1$ fixed and varying values of ρ_1, ρ_2, ρ_3 , and ρ_4 . For settings 10 – 12, we allowed Y_1 and Y_2 to have different marginal error variances, so that the diagonal elements of Σ are equal to (10, 1, 10) and the diagonal elements of Ψ are equal to (10, 10, 10). These are of the form

$$\begin{aligned}\Sigma &= R \begin{pmatrix} 10 & 3.162 & 10 \\ 3.162 & 1 & 3.162 \\ 10 & 3.162 & 10 \end{pmatrix}, \\ R &= \begin{pmatrix} 1 & \rho_1 & 0.6 \\ \rho_1 & 1 & \rho_3 \\ 0.6 & \rho_3 & 1 \end{pmatrix}; \\ \Psi &= \psi \begin{pmatrix} 1 & \rho_2 & 0.6 \\ \rho_2 & 1 & \rho_4 \\ 0.6 & \rho_4 & 1 \end{pmatrix},\end{aligned}$$

where $\psi = 10$. The values of ρ_1, ρ_2, ρ_3 and ρ_4 are all shown in Table 2.

In each dataset we generated systematic missingness in Y_1 and Y_2 according to a missing at random (MAR) mechanism, requiring that no study is missing both variables, and compared all results to the full data set before the missingness was generated. Specifically, we randomly chose studies in which to impose missingness, and then randomly assigned the variable (Y_1 or Y_2) that would be missing. Thus, the missing data mechanism is MAR given study membership. This mechanism mimics the real world situation where the choice of variables to measure made that is within each study reflects the goals of the study, but does not depend on the underlying values of those variables.

We imputed Y_1 and Y_2 with the the goal of correct inference for parameters of the analysis model

$$Y_{i3j} = \beta_0 + \beta_1 Y_{i1j} + \beta_2 Y_{i2j} + z_i b_i + \varepsilon_{ij}; \quad i=1, \dots, m; j=1, \dots, n_i \quad (17)$$

$$b_i \sim N(0, \psi); i=1, \dots, m; \quad \varepsilon_{ij} \sim N(0, \sigma), j=1, \dots, n_i, i=1, \dots, m$$

where the variable Y_3 is fully observed and is the only fixed effect to be used in the imputation models; that is, the joint imputation model is (2) with $\mathbf{X}_j = (1_{n_j} \ Y_{i3})$.

In simulations 1, 2, and 3, we used the random intercept model with $z_i = 1_{n_j}$ so the additional fixed effect $z_i^T Y_{i(-k)}$ is equal to the within-study sum $\sum_{j=1}^{n_i} Y_{i(-k)j}$. The studies used varying numbers of studies and size of studies, which are shown in Table 1. Simulation 1 compares methods in a best-case scenario where sample sizes are moderate within study ($n_j = 30$) and there are many studies ($m = 35$), while Simulation 2 uses a more realistic case of moderate number of studies ($m = 15$) and few without systematic missingness. Simulation 3 uses a moderate number of studies and studies of varying sizes, so that the conditional FCS distributions cannot be compatible with the joint model because their parameters depend on $z_i^T z_i = n_i$. Simulation 4 further investigates the effect of incompatibility by taking z_j to be a vector of independent standard normal random variables, so that b_j is a random slope. In this study, as in Simulation 3, the conditional parameters will vary across studies,

even with equal study sizes. In both Simulations 3 and 4, the imputation models that include $z_i^T Y_{i(-k)}$ in $\mathbf{X}_{ik}^*$ are not strictly compatible with the joint model, but we continue to refer to these as the “compatible” models. We expect that the additional fixed effect makes the imputation model a close approximation to the true compatible model.

In Simulation 5, we assess the influence of the prior hyperparameters on ψ_1^* and ψ_2^* when MICE is used with inverse-Wishart priors in both small and large data settings. We performed this comparison using the same linear generative model $Y_{ij} = \epsilon_{ij} + z_i^T b_i$, but with a single value of Σ and Ψ :

$$\Sigma = \begin{pmatrix} 127 & 100 & 75 \\ 100 & 691 & 65 \\ 75 & 65 & 120 \end{pmatrix}; \Psi = \begin{pmatrix} 28 & 22 & 11 \\ 22 & 40 & 12 \\ 11 & 12 & 26 \end{pmatrix}.$$

These values were chosen to produce a covariance structure similar to that estimated from the TBI dataset discussed in Section 5. In setting A, we again mimicked the TBI dataset and used five studies with sample sizes of 49, 18, 15, 29, and 60. We generated systematic missingness in study 2 (missing Y_1) and study 5 (missing Y_2). Setting B used datasets that were identical in structure and missingness pattern, but contained a much larger number of studies in order to minimize prior influence. We replicated the sample sizes in setting A seven times to get a total of $m = 35$ studies and generated missingness in Y_1 for the studies of size 18 and in Y_2 for the studies of size 60. The analysis model used was the same as for Simulations 1–3, given in (17).

Three inverse-Wishart priors with different degrees of freedom were used to impute missing values for the MICE-IW and MICE-IW-c methods: the default IW(1, 1) prior and the more concentrated IW(1.6, 1) and IW(5, 1) priors. The IW(1.6, 1) prior was chosen because samples from this distribution are distributed similarly to samples of $\psi_{1,2}$ and $\psi_{2,1}$ when the priors on Σ and Ψ are both the default used in the joint distribution, IW(2, $\mathbf{I}_2$) (shown in Figure 2). These prior distribution are displayed in Figure 4.

A total of 200 datasets were constructed for each simulation study. Following [24], we chose $D = 60$ imputations for each method to exceed the number necessary to produce reasonable Monte Carlo error across all of our simulations. Imputations were stored following a burn-in period of 45 iterations. Parameter estimates for the substantive models were combined using Rubin’s rules [2].

4.1. Results

Because many of the simulation results were similar, we present only a subset here. Full results can be found in the supplemental appendix.

Simulations 1 and 2, Balanced Random Intercept—Table 3 shows summaries of parameter estimates and Monte Carlo estimates of standard errors for Simulation 1. We see that, with the exception of the random effect variance ψ , the four imputation methods produce nearly identical estimates and standard errors for all model parameters, and these

are in turn close to the full data estimates. Table 4 shows similar results for Simulation 2; in fact, for all correlation settings in Simulations 1, 2 and 3, only results for ψ differ substantially, and so we focus our presentation on this single parameter.

The first panels of Figure 5 show estimates for the random effect variance ψ in the balanced Simulations 1 and 2, and Table 5 contains these estimates for Simulation 2. All methods seem to perform similarly for settings 1–9 when $m = 35$, but when $m = 15$ there is a slight discrepancy across methods in settings 3, and 6, unusual cases in which the within-study correlation is negative and the between-study correlation is positive. In particular, the MICE-IW methods both seem to be somewhat less efficient in estimating ψ for these settings. This does not appear to be the case for the regular MICE-ML method. An interesting pattern emerges in settings 10–12, where the true error variance of Y_2 is small in comparison to the other parameters: MICE-IW tended to overestimate the random effect variance and its estimates are highly variable. The MICE-IW-c and MICE-ML method are more accurate: including the fixed effect to ensure compatibility appears to alleviate the problem in all three of these settings. These differences may be particularly evident in settings 10 and 11 because the true value of the coefficient associated with the study sums is relatively large in these settings. In settings 10 and 11, we also see a small downward bias in estimating ψ for the joint model imputation.

Simulation 3, Unbalanced Random Intercept—This simulation study highlights the effect of FCS imputation models that require $\psi_{1,2}$, $\beta_{1,2,2}$, $\psi_{2,1}$, and $\beta_{2,1,2}$ to be equal across studies of different sizes. Figure 5 and Table 6 shows that again, the methods performed similarly in estimating ψ except in all but a few settings. As in Simulations 1 and 2, there is a slight inefficiency of MICE-IW in settings 3 and 6, and an upward bias of MICE-IW in settings 10 and 11. Unlike the previous simulations, however, the performance of MICE-ML is notably worse for several cases, producing estimates that are inflated and inefficient for settings 3, 6, 10, and 12. It is not surprising these cases highlight these differences because the conditional parameters in settings 3, 6, 10, and 12 vary considerably as sample size changes, decreasing in magnitude as within-study sample sizes increase. The relatively poor performance of MICE-ML suggests that drawing these parameters from empirical procedures can lead to estimates that are heavily influenced by the smaller studies.

The results of both Simulations 2 and 3 suggest that in some situations, MICE-IW-c can perform notably better than MICE-IW, and so there is at times a benefit to including the study-specific sums in the univariate regression models. Even if the MICE-IW-c model is not truly “compatible” in Simulation 3, its relatively good performance suggests that including the study sums in the imputation model produces an adequate approximation to the true full conditionals. The reduced efficiency of the MICE-ML method in this study suggests that for some unusual covariances, using a proper prior on variance parameters might offer advantageous stabilization.

Simulation 4, Balanced Random Slope—The fourth panel of Figure 5 shows estimates of ψ , the variance of the random slope. All of the MICE methods inflate estimates for settings 10 and 11, but there is no notable difference across these methods. Table 7 details these findings. As shown in Figure 6, we also see a discrepancy across methods in

estimating β_1 and β_2 , the regression coefficients associated with the variables with missing variables. The differences are more apparent in β_1 , the coefficient associated with the variable with higher missingness: both of the MICE-IW methods tend to underestimate the strength of association, particularly the compatible model, and this is seen to some degree in all settings except 1, 4, and 7, the settings where Y_1 and Y_2 are most weakly associated. The joint model and MICE-ML both yield estimates reasonably close to the full data for this parameter.

Simulation 5, Different prior distributions—Figure 7 shows the effect of different choices of prior distributions for both small and large numbers of studies. For the small number of studies in Setting A, we clearly see that the prior distributions can change both the point estimates for the random effects variance and, more dramatically, the variability of such estimates. The effect for the “incompatible” MICE-IW method is most striking, where the estimate and its variance become nearly identical to the approximately unbiased joint model results when the similar inverse-Wishart prior with 1.6 degrees of freedom, is employed. In this case, stronger prior distributions have little effect. However, for the “compatible” version, MICE-IW-c, the stronger priors in the imputation procedure results in dramatically more efficient, though slightly downward biased, estimates of random effect variance. As expected, when the number of studies grows to $m = 35$, comparable to Simulation 1, the choice of prior distribution has little effect. While we did not study the sensitivity to choice of prior distribution for the joint model, the comparability of the joint imputation-based estimates to the estimates with no missing data suggest that the default prior distribution performs relatively well in realistic situations, as in the example presented in Section 5.

5. Example: Traumatic brain injury data

We now apply these methods to data from two multi-site longitudinal studies of children who experience traumatic brain injuries. Previous research has shown that traumatic brain injuries are often associated with psychological and behavioural problems later in life, particularly when the injury is severe [25], and so symptoms observed weeks and months after injury are of interest. Our dataset contains records from two studies: one of preschool-aged children (pTBI) and another of school aged children (TBI) at three sites in Ohio (Cincinnati, Cleveland, and Columbus). Information about these studies can be found in [25] and [26], respectively. While data at several time points are available for some of the participants, we focus on a cross-sectional analysis of the long-term outcomes at 12 months post-injury.

We combine information from these two studies using the multilevel models described above. For the purpose of illustration we treat each site/study combination as a pseudo-study. This approach will both improve algorithmic stability in estimating between-study covariance and better mimic the between-study variability seen in usual IPD meta-analyses. A summary of sample sizes within each pseudo-study is in Table 8. It is reasonable to assume that data within each of these pseudo-studies might be correlated, although the assumption of independent random effects might be less plausible because some pseudo-

studies share the same site or study. While a fixed effect for study and site would alleviate some of this concern, the number of studies precludes this solution.

These datasets illustrate typical missing data issues in IPD meta-analysis, where some measures are taken in both the pTBI and TBI studies, and others are recorded in only one of the two, thus producing systematic missingness. In addition, of course there was also sporadic missingness where some measures are not available due to missed appointments or other individual circumstances, as summarized in Table 8. We used the MICE-IW-c method to easily impute sporadically missing values for the 13 subjects with incomplete data. We do not expect our treatment of this very small amount of missing data to substantially effect our results.

We have chosen two example analyses to specifically illustrate two different types of modeling settings discussed in Section 4. The first analysis focuses on psychological scale outcomes using standard Child Behavior Checklist (CBC) sub-scales, where we generate missingness so as to be able to compare results from the imputation techniques to the ‘true’ estimates with no missing data. The second focuses on measures of socioeconomic status – a particular area of concern for IPD meta-analysis in medical and social science due to immense variation across studies in the choice of which measure to use.

Example 1: CBC scores

In our first example, we consider the relationship between the Brief Symptom Inventory (BSI) overall score [27] and the internalizing and externalizing subscales of the Child Behavior Checklist (CBC) [28, 29]. The BSI provides a summary measure for general psychological wellness. The CBC internalizing score captures indications of internalizing problems such as anxiety or depression, and the externalizing score captures indications of externalizing problems such as aggressive behavior. Higher CBC scores indicate greater prevalence of problem behaviors. Both instruments and all three summaries are well-established, often used in psychological studies, and potentially related to long-term behavioral effects of head injury. In these two studies, both questionnaires were completed by each child’s primary caregiver. Although the versions of the CBC used among the pTBI differed from the TBI, due to the ages of the children, both versions provide a summary score related to the same underlying constructs. Further, the CBC subscale scores are standardized so that systematic differences between the versions should be minimized. Finally, we follow usual practice by including in the analysis both demographic variables (age, sex, race, maternal education) and a binary group variable that indicates whether the injury was “severe” or “mild/moderate”. Summaries of these variables are shown in Table 9.

Our goal in this first example is to mimic a case where a subset of different psychological scale measures are used in different studies. Both CBC subscale scores were fully observed in both the pTBI and TBI studies, so we generated systematic missingness by deleting the externalizing subscale in pseudo-study 2 and the internalizing subscale in pseudostudy 5. Because these two scores are derived from the same checklist, it is unlikely in practice that one would be systematically missing; nonetheless, it is plausible that systematically missing variables may have a similar association, such as scores from two different indices of injury severity or methods of summarizing the same checklist.

We imputed missing values according to the four methods described in Section 2. For the joint imputation model in Equation 1, we specified the outcome Y_i to be the CBC internalizing and externalizing scores and the covariate matrix $\mathbf{X}_i$ to comprise the injury group, sex, age, race, maternal education, and BSI score. For the MICE-IW imputations based on Equation 5, we specified the covariate matrix $\mathbf{X}_{(1)i}^*$ to comprise $\mathbf{X}_i$ and the CBC externalizing score, $Y_{(2)i}$, and similarly the covariate matrix $\mathbf{X}_{(2)i}^*$ to comprise $\mathbf{X}_i$ and the CBC internalizing score, $Y_{(1)i}$. The MICE-IW-c and MICE-ML imputations add the study-specific averages of the appropriate CBC score, $\bar{Y}_{(k)j}$ to the fixed effect covariates of the MICE-IW model. For all imputation models, the random effects covariates, Z_i and Z_i^* , are simply study indicators corresponding to random intercepts models. We analyzed completed datasets using a random intercepts regression of the BSI score on the CBC subscale scores and the baseline variables listed in Table 10. Summaries of parameter estimates are also shown in Table 10.

The parameter estimates suggest a positive association between both the internalizing and externalizing CBC scores and BSI score. Relative to the full data, most of the imputation methods produce coefficients that are slightly overestimated for the internalizing scores and underestimated for the externalizing scores. The exception is the MICE-ML method, which slightly overestimates the CBC externalizing score. Several other variables' coefficients (e.g., sex and race) differ more substantially between full data and imputations. However, all of these differences are small relative to the estimated standard errors. Consistent with the simulation results, the four imputation methods vary in their estimates of the random effect variance. All methods over-estimate the between-study variability by at least 50%, but MICE-IW-c produces an estimate nearly 150% larger than the full data estimate. The standard error of this estimate is also substantially larger than for the other estimates. We note that this relative instability of estimates using MICE-IW-c can be alleviated if the "compatible" study means are omitted from the imputation models for sporadically missing values, reducing the number of parameters that must be estimated. We expect that in most cases, there is little to be gained by requiring compatible imputation models for sporadically missing variables. The four imputation methods were in agreement for the error variance parameter, which is much larger than the random effect; the model fit suggests that the between-study variability of this example is relatively small in magnitude.

Example 2

Our second example considers imputing socioeconomic status (SES) variables for a prognostic model of CBC internalizing score. In particular, we consider imputing the Hollingshead index, a measure of socioeconomic status observed only in the school-aged children of the TBI study (pseudo-studies 4 and 5). While many psychological studies rely on census tract median income or maternal education to capture SES, the Hollingshead index (and related measures, such as the Duncan score) also incorporate occupational information. Thus, the Hollingshead index may be a more accurate measure of SES, but is much more difficult to measure and construct in practice. SES was measured using other metrics in both the pTBI and TBI studies; each report maternal education and a normed composite, which we call the "z-ses" score. (Although the z-ses score is based on different

variables across the studies, the measures are normed to the sample and thus should be comparable for this example.) To mimic the motivating example of two variables with systematic missingness, we also imposed artificial missingness on z-ses in pseudo-study 4.

We again imputed missing values according to the four methods described in Section 2. For the joint imputation model in Equation 1, we specified the outcome Y_i to be the Hollingshead index and the z-ses composite measure and the covariate matrix $\mathbf{X}_i$ to comprise the injury group, sex, age, race, and maternal education. For the MICE-IW imputations based on Equation 5, we specified the covariate matrix $\mathbf{X}_{(1)i}^*$ to comprise $\mathbf{X}_i$ and z-ses, $Y_{(2)i}$, and similarly the covariate matrix $\mathbf{X}_{(2)i}^*$ to comprise $\mathbf{X}_i$ and the Hollingshead index, $Y_{(1)i}$. The MICE-IW-c imputation adds the study-specific averages of the appropriate scores, $\bar{Y}_{(k)i}$ to the fixed effect covariates of the MICE-IW model. The MICE-ML imputation adds only the study-specific Hollingshead average to the imputation model for z-ses; because only one study includes both the Hollingshead and z-ses measures, adding the study-specific z-ses averages to the Hollingshead imputation would have resulted in unstable maximum likelihood estimates in the imputation algorithm. The need for this adjustment highlights a computational limitation of the MICE-ML approach for sparse overlap in fully observed parameters across studies. Again, for all imputation models, the random effects covariates, Z_i and Z_i^* , are simply study indicators corresponding to appropriate random intercepts models. Our target model for analysis is a regression of CBC internalizing score on the Hollingshead score and the baseline variables listed in Table 11. The parameter estimates based on the four imputation procedures are also shown in Table 11. Because we have no “true” complete data analysis with which to compare our results, this table also includes the estimated fraction of missing information for each coefficient [2].

The baseline indicator of injury group has the strongest relationship with the CBC internalizing score, with the mild/moderate group showing lower prevalence of behavioral problems than the severe injury group. The coefficient associated with Hollingshead index suggests a negative association between socioeconomic status and long-term behavioral effects of TBI, but this coefficient is highly variable due to the high missingness, as indicated by estimated fractions of missing information near 40%. As in Example 1, and expected from our simulation study, most of the coefficient estimates are fairly similar across imputation methods. For this model, the data are considerably more variable than in Example 1. Estimates of both error variance and random effect variance are large, for example 110 and 4 for the joint imputation results, respectively. Again, the magnitude of between-study variability is small compared to the within-study variability. While the error variance estimates are relatively similar across the imputation methods, the estimates of the random effect variance differ considerably. The joint method produces the smallest estimate. Based on the simulation results, we expect this estimate to be closest to the truth. The MICE methods all estimate the random effect variance to be larger, with MICE-IW and MICE-IW-c indicating variances close to 8 or 9. Our modified MICE-ML procedure results in a substantially larger estimate of the random effect variance of 20, with an accompanying large standard error over 50. Because the MICE-ML algorithm relies on ML estimates, we are not surprised that it performs relatively poorly in this sparse imputation example.

6. Discussion

In this paper we compared joint model imputation and MICE in terms of their theoretical equivalence and empirical performance for the multilevel multivariate normal model with systematically missing data in the structure commonly seen for IPD meta-analysis. We showed in Section 3 that the two methods may not be theoretically equivalent in most MICE models; namely, the equivalent full conditionals require a study-specific term in the regression function, and some of the conditional parameters are typically specific to a study even if the joint model specifies homogeneous parameters. These features, in particular the second, mean that it is not realistic to use a fully compatible model for systematically missing covariates. Nonetheless, simulations in Section 4 suggest that in these examples, similar to findings in [9], incompatible conditional FCS models will often still lead to reasonable imputations; analyses based on imputations from joint and FCS procedures do not tend to differ substantially in many common settings. When there are differences, it is in estimating the variance of the random effect. Including the sum for a compatible (or approximately compatible) imputation model can often prevent such difficulties. Our work also suggests that when the conditional parameters vary considerably across studies, use of prior information can improve performance of FCS methods.

Because both models account for variables' multilevel structure, Enders, Mistler and Keller conclude that both FCS and joint model imputation are appropriate for multilevel random intercepts models [14]. Although we agree that FCS is typically able to produce good imputations, our work reveals some features of the full conditional models that might on occasion lead to biased or inefficient estimators. These problems tend to be mild, but might be exacerbated by small samples sizes or unequal error and random effect variances in the presence of systematic missingness. In addition, we found that joint model and FCS methods can diverge in small samples due to the influence of priors. Standard "weakly-informative" inverse-Wishart priors on the joint parameterization, for example, translate to concentrated distributions on the conditional parameterization, and the same priors applied directly to the conditional parameterization often lead to predictive distributions with heavier tails. The results of Simulation 5 suggest that a data-driven approach with no prior influence, as in the MICE-ML method, can at times mitigate this influence. Nonetheless, in the same situations when prior influence is large, data-based hyperparameters may also be unreliable. Our work regarding these priors is not comprehensive, however, and other families of distributions might result in improved performance. Overall, because the prior distribution specification can substantially effect the analytical results, robust sensitivity analyses are particularly important for small collections of studies. Comparing results across a range of prior distributions does not reveal the 'best' prior distribution for any particular analysis, but can convey the level of caution one should take in making inference based on any one chosen prior distribution.

Existing work has emphasized the importance of accounting for data's multilevel structure when accounting for missingness, and multiple imputation using a chained equations approach is often the most flexible and user-friendly method. For example, an approach by The Fibrinogen Studies Collaboration addresses systematic missingness by directly modelling the relationship between estimates from fully observed studies and those with

missing values, but requires likelihood calculations that are not easily adapted to differing patterns of missingness [30]. FCS, on the other hand, can easily adapt to a variety of missingness patterns and typically requires specification of only simple regression models. Our results suggest that the multiple imputation FCS alternatives will typically perform adequately for a random intercepts model. Although the random slopes case was not our primary focus, the results of simulation 4 suggest that joint model imputation is slightly better suited to capture all associations for this model when missingness is high.

Here we considered only the multivariate normal model where both methods are easily implemented and derivations are analytically tractable. In practice, FCS is often used in cases where an appropriate joint model is less straightforward, such as when missing values occur in both continuous and categorical variables, where a joint model is much less easily implemented. The work of Enders, Mistler and Keller [14] and Kline, Andridge and Kaizar [17] suggest that a latent variable imputation scheme may be an effective method of utilizing joint normal imputation methods to impute categorical variables, and the authors showed promising improvement over a chained equations approach based on rounding. Their work did not address systematic missingness, however. For systematic missingness, Jolani, et al provides a framework for FCS imputation in a method analogous to those discussed here, but within a generalized linear model framework that is appropriate for binary and categorical variables [13]. Future work might compare the latent variable joint imputation to this approach to non-normally distributed data.

Even though our comparisons in the continuous case are not directly applicable to model non-normally distributed variables, we believe that our results here, nonetheless, do reveal properties of fully conditional models that are applicable to other problems. First, it is likely that homogeneity assumptions in the fully conditional model will be misspecified for multilevel models, which may negatively affect imputations in some cases. Second, prior influence can be strong on the parameters corresponding to the higher level of the hierarchy.

In sum, we have found that for many cases of multivariate normal hierarchical data with systematic missingness, the joint or conditional specification of imputation procedures in itself has little impact on analytical results. However, especially in cases with few groups or studies consistent with typical IPD meta-analyses, the choice of prior distributions substantially affects the results. As such, ‘default’ prior distributions should be used only with caution. In absence of other guidelines, we believe that using joint model imputation with default priors, or FCS priors that are similar to the joint priors, will often work well. Nonetheless, we recommend that researchers carefully consider prior specifications and their implications regardless of the chosen imputation method, and explore the importance of this choice via sensitivity analyses.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors thank Drs. Terry Stancin, H. Gerry Taylor, Shari Wade, and Keith Yeates for providing data and assistance related to the example.

Contract/grant sponsor: This work was supported in part by the Eunice Kennedy Shriver National Institute of Child Health and Human Development under award number R03HD080033. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

References

1. Resche-Rigon M, White IR, Bartlett JW, Peters SAE, Thompson SG. Multiple imputation for handling systematically missing confounders in meta-analysis of individual participant data. *Statistics in Medicine*. 2013; 32:4890–4905. DOI: 10.1002/sim.5894 [PubMed: 23857554]
2. Rubin, D., Little, RJA. *Statistical Analysis with Missing Data*. John Wiley and Sons; Hoboken, New Jersey: 2002.
3. Thompson SG, Higgins JPT. How should meta-regression analyses be undertaken and interpreted? *Statistics in Medicine*. 2002; 21:1559–1573. DOI: 10.1002/sim.1187 [PubMed: 12111920]
4. Andridge RR. Quantifying the impact of fixed effects modeling of clusters in multiple imputation for cluster randomized trials. *Biometrical Journal*. 2011; 53:57–74. DOI: 10.1002/bimj.201000140 [PubMed: 21259309]
5. Siddique J, Reiter JP, Brincks A, Gibbons RD, Crespi CM, Brown CH. Multiple imputation for harmonizing longitudinal non-commensurate measures in individual participant data meta-analysis. *Statistics in Medicine*. 2015; 34(26):3399–3414. URL <http://dx.doi.org/10.1002/sim.6562>, sim. 6562. DOI: 10.1002/sim.6562 [PubMed: 26095855]
6. Schafer, JL. Technical Report 97-04. The Pennsylvania State University; 1997. Imputation of missing covariates under a multivariate linear mixed model.
7. Schafer JL, Yucel RM. Computational strategies for multivariate linear mixed-effects models with missing values. *Journal of Computational and Graphical Statistics*. 2002; 11(2):437–457. DOI: 10.1198/106186002760180608
8. van Buuren S, Groothuis-Oudshoorn K. mice: Multivariate imputation by chained equations in R. *Journal of Statistical Software*. 2011; 45(1):1–67. DOI: 10.18637/jss.v045.i03
9. van Buuren S, Brand JPL, Groothuis-Oudshoorn CGM, Rubin DB. Fully conditional specification in multivariate imputation. *Journal of Statistical Computation and Simulation*. 2006; 76(12):1049–1064. DOI: 10.1177/096228020607446
10. Arnold B, Press SJ. Compatible conditional distributions. *Journal of the American Statistical Association*. 1989; 84(405):152–156. DOI: 10.1080/01621459.1989.10478750
11. Liu J, Gelman A, Hill J, Su YS, Kropko J. Stationary distribution of iterative imputations. *Biometrika*. 2013; :1–19. DOI: 10.1093/biomet/ast044
12. Siddique, J., de Chavez, PJ., Howe, G., Cruden, G., Brown, CH. Limitations in using multiple imputation to harmonize individual participant data for meta-analysis; *Prevention Science*. 2017. p. 1-14. URL <http://dx.doi.org/10.1007/s11121-017-0760-x>
13. Jolani S, Debray TPA, Koffijberg H, van Buuren S, Moons KGM. Imputation of systematically missing predictors in an individual participant data meta-analysis: a generalized approach using mice. *Statistics in Medicine*. 2015; 34:1841–1863. DOI: 10.1002/sim.6451 [PubMed: 25663182]
14. Enders CK, Mistler SA, Keller BT. Multilevel multiple imputation: A reiew and evaluation of join modeling and chained equations imputation. *Psychological Methods*. 2015; doi: 10.1037/met0000063
15. Yucel R. Multiple imputation inference for multivariate multilevel continuous data with ignorable non-response. *Philosophical Transactions of the Royal Society A*. 2008; 366(1874):2389–2403. DOI: 10.1098/rsta.2008.0038
16. Kline D, Andridge R, Kaizar E. Comparing multiple imputation methods for systematically missing subject-level data. *Research Synthesis Methods*. 2015; doi: 10.1002/jrsm.1192
17. Kline, D., Andridge, R., Kaizar, E. PhD Thesis. The Ohio State University; 2016. Systematically missing subject-level data in longitudinal research synthesis.

18. Zhao JH, Schafer JL. pan: Multiple imputation for multivariate panel or clustered data 2016. R package version 1.4.
19. Hughes R, White IR, Seaman SR, Carpenter JR, Tilling K, Sterne JAC. Joint modelling rationale for chained equations. BMC Medical Research Methodology. 2014; 14doi: 10.1186/1471-2288-14-28
20. Arnold BC, Castillo E, Sarabia JM. Conditionally specified distributions: An introduction. Statistical Science. 2001; 16(3):249–274. DOI: 10.1214/ss/1009213728
21. Bartlett J, Seaman SR, White IR, Carpenter JR. Multiple imputation of covariates by fully conditional specification: Accommodating the substantive model. Statistical Methods in Medical Research. 2015; 24(4):462–487. DOI: 10.1177/0962280214521348 [PubMed: 24525487]
22. Carpenter, J., Kenward, M. Multiple Imputation and its Application. John Wiley and Sons; 2013.
23. Demidenko, E. Mixed Models. John Wiley and Sons; Hoboken New Jersey: 2004.
24. White IR, Royston P, Wood AM. Multiple imputation using chained equations: Issues and guidance for practice. Statistics in Medicine. 2011; 30(4):377–399. [PubMed: 21225900]
25. Chapman LA, Wade SL, Walz NC, Taylor HG, Stancin T, Yeates KO. Clinically significant behavior problems during the initial 18 months following early childhood traumatic brain injury. Journal of Rehabilitation Psychology. 2010; 55:48–57. DOI: 10.1037/a0018418 [PubMed: 20175634]
26. Fay TB, Yeates KO, Wade SL, Drotar D, Stancin T, Taylor HG. Predicting longitudinal patterns of functional deficits in children with traumatic brain injury. Neuropsychology. 2009; 23:271–282. [PubMed: 19413442]
27. Derogatis L, Melisaratos N. The brief symptom inventory: an introductory report. Psychological Medicine. 1983; 13:595–605. [PubMed: 6622612]
28. Achenbach, T., Rescorla, L. Manual for the ASEBA Preschool Forms and Profiles. University of Vermont Department of Psychiatry; Burlington, VT: 2001.
29. Achenbach, T., Rescorla, L. Manual for the ASEBA School-Age Forms and Profiles. University of Vermont Department of Psychiatry; Burlington, VT: 2001.
30. The Fibrinogen Studies Collaboration. Systematically missing confounders in individual participant data meta-analysis of observational cohort studies. Statistics in Medicine. 2009; 28(8): 1218–1237. DOI: 10.1002/sim.3540 [PubMed: 19222087]

Appendix

Notation

Data Structure		Constant Objects
$i = 1, \dots, m$	Index for 'level-1' group, or study	$\mathbf{1}_n$ Column vector of ones, length n
$j = 1, \dots, n_i$	Index for 'level-2' individuals, or participants within study i	$\mathbf{I}_n$ Square identity matrix of dimension n
$k = 1, \dots, r$	Index for variable	$\mathbf{J}_n$ Square matrix of dimension n of all ones
$\mathbf{Y}^{mis}$	The portion of $\{\mathbf{Y}_i\}_{i=1,\dots,m}$ where the values are missing	
$\mathbf{Y}^{obs}$	The portion of $\{\mathbf{Y}_i\}_{i=1,\dots,m}$ where the values are observed	
$\mathcal{M}_k \subset \{1, \dots, m\}$	Collection of indices for studies where variable k is systematically missing	

Objects Used in Models			
Joint Imputation Model for $\mathbf{Y}$ (Equation (1))		Full Conditional Specification (FCS) Imputation Models for $\mathbf{Y}_{(k)}$ (Equations (4))	
Likelihood or Conditional Likelihood Functions			
$\mathbf{Y}_i = (Y_{i1}, \dots, Y_{ir}) \in R^{n_i \times r}$	Matrix of r partially observed variables for each of n_i individuals participating in study i	$\mathbf{Y}_{(k)}$	The portion of $\{\mathbf{Y}_i\}_{i=1, \dots, m}$ that only includes the k th variable
		$\mathbf{Y}_{(-k)}$	$\{\mathbf{Y}_i\}_{i=1, \dots, m}$ with the k th variable removed
$\mathbf{X}_i \in R^{n_i \times p}$	Matrix of p fully observed fixed effects covariates for each of n_i individuals participating in study i	$\mathbf{X}_{ik}^* \in R^{n_i \times p^*}$	Matrix of p^* fixed effect covariates; possibly includes $\mathbf{Y}_{i(-k)}$ or transformations of $\mathbf{Y}_{i(-k)}$.
$\mathbf{Z}_i \in R^{n_i \times q}$	Matrix of q fully observed variables for each of n_i individuals participating in study i ; typically associated with random effects.	$\mathbf{Z}_{ik}^* \in R^{n_i \times q^*}$	Matrix of q^* random effect covariates for the conditional model of response variable $\mathbf{Y}_{(k)}$; possibly includes $\mathbf{Y}_{i(-k)}$ or transformations of $\mathbf{Y}_{i(-k)}$.
$\mathbf{e}_i = (\mathbf{e}_{i1}, \dots, \mathbf{e}_{in_i})' \in R^{n_i \times r}$	Matrix of errors, where each row represents the vector of errors for individual j in group i .		
$\boldsymbol{\beta} \in R^{p \times r}$	Matrix of fixed effects	$\boldsymbol{\beta}_k^* \in R^{p^*}$	Vector of fixed effects
$\mathbf{B}_i \in R^{q \times r}$	Matrix of random effects for study i	$\mathbf{B}_{(k)}^* \in R^{m \times q^*}$	Random effects matrix
$\boldsymbol{\Sigma} \in R^{r \times r}$	Variance-covariance matrix for the individual-specific error vector $\mathbf{e}_{ij}$	σ_k^*	Scalar error variance
$\boldsymbol{\Psi} = \{\psi_{ab}\}_{a=1, \dots, r; b=1, \dots, r; q} \in R^{q \times r \times q}$	Variance-covariance matrix for the stacked vector of random effects, $\text{vec}(\mathbf{B}_i)$	$\boldsymbol{\psi}_k^* \in R^{q^* \times q^*}$	Random effect variance
$\boldsymbol{\theta} = \{\boldsymbol{\beta}, \boldsymbol{\Sigma}, \boldsymbol{\Psi}\}$	All parameters for the likelihood of the variables with missing values, $\mathbf{Y}$	$\boldsymbol{\theta}_k^* = \{\boldsymbol{\beta}_k^*, \sigma_k^*, \boldsymbol{\psi}_k^*\}$	Vector of parameters for the fully conditional specified likelihood for variable with missing values $\mathbf{Y}_{(k)}$
Prior Distributions			
π'	Prior distribution for $\boldsymbol{\theta}$ in the joint imputation model	π_k^{FCS}	Prior distribution for the parameters $\boldsymbol{\phi}_k$ and latent random effects $\mathbf{B}_{(k)}^*$
		ν_k	Degrees of freedom for inverse-Wishart prior on variance parameters
		V_k	Scale for inverse-Wishart prior on variance parameters
Conditional Distributions for Sampling Algorithm			
P	Full conditional distribution of the parameter or latent variable • in the Gibbs sampler derived		

Objects Used in Models			
Joint Imputation Model for Y (Equation (1))		Full Conditional Specification (FCS) Imputation Models for $Y_{(k)}$ (Equations (4))	
$P_k(Y_{(k)} Y_{(-k)}, \theta_k)$	from the joint imputation model and prior distribution π^J Full conditional distribution for the parameters θ_k and latent random effects $B_{(k)}^*$ as derived from the joint likelihood and prior π_k^J	$\tilde{g}_k$	Pseudo-full conditional distribution for the parameters θ_k^* and latent random effects $B_{(k)}^*$ as derived from the fully conditional specified π_k^{FCS} likelihood and prior
θ_k	Parameters for the full conditional distribution $P_k^Y(Y_{(k)} Y_{(-k)}, \theta_k)$	g_k	Pseudo-full conditional distribution for the missing values for variable k , Y_k^{mis} as derived from the fully conditional specified likelihood

Figure 1.

True values of $\beta_{1,2,2}^i$ and $\psi_{1,2}^i$ when $\psi_{11} = 27$, $\psi_{22} = 33$, $\sigma_{11} = 122$, $\sigma_{22} = 147$, $\sigma_{12} = 100$.

Lines shown depict varying values of $\rho = \frac{\psi_{12}}{\sqrt{\psi_{11}\psi_{22}}}$.

Figure 2. Marginal density estimates from samples of θ_2^* . Improper flat prior densities for $\beta_{2,1}$ parameters under the FCS model are not shown.

Figure 3.

Quantile quantile plots for the joint and MICE-IW predictive distributions for selected individuals Y_{ij} ; Row 1: $n=10$, $m=5$; Row 2: $n=30$, $m=5$; Row 3: $n=10$, $m=30$; Row 4: $n=30$, $m=30$. Ψ takes the same values as in Figure 1 with $\rho = .85$. Σ is the matrix used in Figure 1 with correlation 0.65, but scaled by $1/1.75$.

Figure 4.
Densities of the three priors used for the random effect variances in Simulation 5.

Figure 5.

Estimates of random effect variance for Simulations 1–4. Lines represent two standard errors from the point estimate. Note the different scales across graphs.

Figure 6.
Estimates of β_1 and β_2 for simulation study 4.

Simulation 5

Figure 7.
Estimates of ψ for simulation study 5.

Table 1

Description of the simulation studies

Simulation	Study sizes	m	$\sum_i^m n_i$	Studies missing Y_1, Y_2
1. Balanced, large m	$n_i = 30$ for all i	35	1050	8, 9
2. Balanced, small m	$n_i = 30$ for all i	15	450	8, 3
3. Unbalanced, small m	n_i varying in $[15, 150]$	15	450	8, 3
4. Balanced, small m , random slope on $z_i \sim N(0, 1)$	$n_i = 30$ for all i	15	450	8, 3
5a. Unbalanced, small m	49, 18, 15, 29, 60	5	171	1, 1
5b. Unbalanced, small m	Replacates of 5a	35	1197	7, 7

Table 2

Correlation settings used for simulation studies 1–4

Setting	1	2	3	4	5	6	7	8	9	10	11	12
ρ_1	0.15	0.75	-0.55	0.15	0.75	-0.55	0.15	0.75	-0.55	-0.55	-0.55	0.75
ρ_2	0.15	0.15	0.15	0.75	0.75	0.75	-0.55	-0.55	-0.55	0.75	-0.55	0.75
ρ_3	0.40	0.40	0.40	0.40	0.40	0.40	0.40	0.40	0.40	-0.40	-0.40	0.40
ρ_4	0.40	0.40	0.40	0.40	0.40	0.40	0.40	0.40	0.40	0.40	-0.40	0.40

Table 3
Simulation 1, correlation setting 6 mean parameter estimates and Monte Carlo (standard errors)

	β_0	β_1	β_2	σ	ψ
Full data	-0.039 (0.32)	1.164 (0.02)	1.035 (0.02)	3.753 (0.17)	3.440 (0.46)
Joint	-0.042 (0.35)	1.164 (0.03)	1.036 (0.03)	3.741 (0.23)	3.369 (0.73)
MICE-IW	-0.055 (0.40)	1.163 (0.03)	1.035 (0.03)	3.759 (0.23)	3.167 (0.84)
MICE-IW-c	-0.043 (0.35)	1.164 (0.03)	1.035 (0.03)	3.761 (0.23)	3.206 (0.73)
MICE-ML	-0.046 (0.35)	1.164 (0.03)	1.035 (0.03)	3.756 (0.23)	3.152 (0.71)

Table 4
Simulation 2, correlation setting 6 mean parameter estimates and Monte Carlo (standard errors)

	β_0	β_1	β_2	σ	ψ
Full data	-0.047 (0.48)	1.166 (0.04)	1.036 (0.04)	3.783 (0.26)	3.420 (0.69)
Joint	-0.049 (0.59)	1.168 (0.06)	1.036 (0.06)	3.791 (0.47)	3.155 (1.59)
MICE-IW	-0.034 (0.80)	1.164 (0.06)	1.031 (0.06)	3.837 (0.48)	3.264 (2.91)
MICE-IW-c	-0.064 (0.68)	1.166 (0.06)	1.033 (0.06)	3.830 (0.48)	3.514 (2.35)
MICE-ML	-0.062 (0.60)	1.169 (0.06)	1.036 (0.05)	3.808 (0.46)	2.993 (1.51)

Table 5

Estimate of ψ ; Simulation 2

	Full data	Joint	MICE-IW	MICE-IW-c	MICE-ML
1	1.041 (0.38)	1.057 (0.48)	1.075 (0.48)	1.131 (0.53)	1.081 (0.48)
2	1.270 (0.42)	1.326 (0.52)	1.318 (0.53)	1.373 (0.55)	1.340 (0.52)
3	2.039 (0.53)	2.326 (1.38)	2.094 (1.88)	2.584 (1.91)	2.139 (1.25)
4	1.218 (0.41)	1.171 (0.49)	1.123 (0.48)	1.144 (0.49)	1.100 (0.46)
5	1.133 (0.39)	1.172 (0.46)	1.158 (0.45)	1.203 (0.48)	1.146 (0.44)
6	3.420 (0.69)	3.155 (1.59)	3.264 (2.91)	3.514 (2.35)	2.993 (1.51)
7	0.715 (0.31)	0.884 (0.43)	0.960 (0.45)	1.008 (0.48)	0.954 (0.43)
8	1.261 (0.41)	1.268 (0.51)	1.330 (0.57)	1.343 (0.54)	1.313 (0.51)
9	0.369 (0.23)	1.202 (1.00)	0.830 (0.86)	1.333 (1.35)	0.773 (0.68)
10	12.287 (1.28)	10.678 (2.59)	15.063 (4.05)	12.274 (2.55)	12.155 (2.42)
11	11.619 (1.25)	10.469 (2.94)	14.514 (4.84)	12.188 (2.90)	12.088 (2.70)
12	11.598 (1.25)	11.607 (2.78)	12.749 (3.09)	11.890 (2.53)	11.819 (2.37)

Table 6Estimates of ψ ; Simulation 3

	Full data	Joint	MICE-IW	MICE-IW-c	MICE-ML
1	1.040 (0.38)	1.070 (0.46)	1.076 (0.46)	1.126 (0.50)	1.298 (0.65)
2	1.214 (0.40)	1.271 (0.50)	1.253 (0.52)	1.301 (0.53)	1.305 (0.55)
3	2.133 (0.54)	2.499 (1.40)	2.335 (2.16)	2.704 (2.07)	4.587 (4.03)
4	1.245 (0.41)	1.150 (0.48)	1.096 (0.47)	1.129 (0.50)	1.364 (0.70)
5	1.201 (0.40)	1.258 (0.48)	1.233 (0.46)	1.273 (0.49)	1.267 (0.49)
6	3.642 (0.70)	3.221 (1.61)	3.362 (2.94)	3.371 (2.25)	7.270 (5.46)
7	0.752 (0.32)	0.933 (0.43)	1.014 (0.45)	1.059 (0.48)	1.276 (0.69)
8	1.310 (0.42)	1.305 (0.52)	1.305 (0.56)	1.344 (0.55)	1.363 (0.60)
9	0.387 (0.23)	1.198 (1.00)	0.815 (0.87)	1.219 (1.20)	1.596 (1.93)
10	12.537 (1.29)	11.211 (3.06)	16.351 (4.82)	12.837 (3.09)	17.239 (6.81)
11	12.343 (1.28)	11.583 (3.51)	15.025 (4.95)	13.045 (3.22)	14.337 (4.22)
12	11.784 (1.25)	12.184 (3.18)	13.132 (3.30)	12.147 (2.88)	13.887 (4.74)

Table 7

Estimates of ψ ; Simulation 4

	Full data	Joint	MICE-IW	MICE-IW-c	MICE-ML
1	1.096 (1.45)	1.121 (1.50)	1.200 (1.56)	1.204 (1.56)	1.165 (1.53)
2	1.203 (1.51)	1.300 (1.62)	1.418 (1.72)	1.397 (1.71)	1.417 (1.72)
3	2.091 (2.02)	2.240 (2.39)	2.431 (2.96)	2.490 (3.18)	2.337 (2.79)
4	1.269 (1.56)	1.204 (1.54)	1.219 (1.56)	1.225 (1.56)	1.198 (1.54)
5	1.114 (1.44)	1.217 (1.56)	1.284 (1.61)	1.276 (1.61)	1.292 (1.62)
6	3.609 (2.65)	3.154 (2.78)	3.971 (4.53)	3.841 (3.98)	3.523 (3.60)
7	0.754 (1.19)	0.924 (1.35)	1.067 (1.47)	1.065 (1.47)	1.029 (1.44)
8	1.251 (1.53)	1.350 (1.65)	1.595 (1.85)	1.579 (1.84)	1.560 (1.82)
9	0.369 (0.84)	1.261 (1.77)	1.145 (1.86)	1.148 (1.89)	1.214 (1.88)
10	11.827 (4.78)	10.221 (5.31)	20.685 (9.43)	20.778 (9.20)	20.073 (8.35)
11	11.360 (4.70)	9.812 (5.19)	16.268 (8.70)	16.145 (8.53)	15.765 (7.72)
12	11.667 (4.76)	10.546 (5.13)	13.745 (6.69)	13.785 (6.53)	13.039 (5.90)

Table 8

Sample sizes and numbers of missing values for each pseudo-study at the 12-month followup visit. Boldface indicated systematic missingness.

Study	pTBI	pTBI	pTBI	TBI	TBI
Site	Site 1	Site 2	Site 3	Site 2	Site 3
Pseudo-Study Indicator	1	2	3	4	5
Sample Size					
	51	21	16	32	64
	184				
Number of missing values for each variable					
Injury Group	0	0	0	0	0
Sex	0	0	0	0	0
Age	0	0	0	0	0
Race	0	0	0	0	0
Maternal Education	0	0	0	2	1
CBC Internalizing	0	3	0	0	0
CBC Externalizing	0	3	0	0	0
BSI Score	2	1	1	1	3
Hollingshead	51	21	16	0	0
z-ses	0	0	0	0	0

Table 9

Proportions or means and (standard deviations) of variables by pseudo-study

Pseudo-Study	1	2	3	4	5
Group (1 = mild/moderate injury)	0.863	0.571	0.688	0.500	0.578
Sex (1 = male)	0.588	0.714	0.438	0.750	0.734
Age	6.006 (1.08)	6.129 (1.31)	6.282 (1.16)	10.824 (2.18)	10.843 (1.95)
Race (1 = Caucasian)	0.765	0.714	0.500	1.000	0.703
Maternal Education	4.412 (1.28)	4.238 (1.00)	4.750 (1.06)	4.167 (1.02)	4.683 (1.16)
CBC Internalizing	53.755 (10.40)	53.200 (11.54)	48.000 (11.58)	56.516 (10.22)	53.230 (11.41)
CBC Externalizing	48.392 (9.46)	53.667 (14.77)	49.875 (11.82)	56.594 (11.05)	52.906 (10.70)
BSI Score	53.569 (12.26)	55.111 (18.77)	48.062 (10.27)	55.188 (11.88)	52.703 (12.12)
Hollingshead	NA (NA)	NA (NA)	NA (NA)	32.281 (12.88)	36.125 (15.55)
z-ses	-0.237 (1.00)	-0.326 (0.88)	-0.087 (1.12)	-0.013 (0.87)	0.190 (1.07)

Table 10

Parameter estimates and (standard errors) for Example 1 analysis of BSI score on listed covariates.

	Full data	Joint	MICE-IW	MICE-IW (c)	MICE-ML
Intercept	26.001 (5.79)	24.443 (5.89)	24.288 (6.04)	24.391 (6.31)	23.101 (6.49)
CBC Internalizing	0.225 (0.08)	0.287 (0.09)	0.260 (0.10)	0.290 (0.10)	0.255 (0.09)
CBC Externalizing	0.302 (0.07)	0.278 (0.08)	0.294 (0.09)	0.268 (0.09)	0.317 (0.08)
Group (1 = mild/moderate injury)	1.084 (1.45)	1.339 (1.50)	0.960 (1.50)	1.073 (1.51)	1.214 (1.51)
Sex (1=male)	-0.200 (1.46)	-0.266 (1.52)	-0.199 (1.53)	-0.382 (1.54)	0.175 (1.60)
Age	0.407 (0.26)	0.337 (0.29)	0.402 (0.29)	0.393 (0.30)	0.380 (0.31)
Race (1=Caucasian)	-1.518 (1.62)	-1.694 (1.71)	-1.736 (1.70)	-1.882 (1.75)	-1.671 (1.72)
Maternal Education	-0.759 (0.63)	-0.706 (0.67)	-0.602 (0.66)	-0.629 (0.70)	-0.680 (0.70)
Error variance	81.389 (8.60)	76.884 (8.62)	76.656 (8.84)	77.716 (9.11)	79.261 (9.52)
Random effect variance	0.919 (0.53)	1.415 (1.49)	1.421 (1.44)	2.186 (2.59)	1.424 (1.77)

Table 11

Parameter estimates, (standard errors), and [estimated fraction of missing information] for Example 2 regression of CBC internalizing score on listed covariates.

	Joint	MICE-IW	MICE-IW (c)	MICE-ML
Intercept	57.089 (5.13) [0.118]	57.810 (5.50) [0.234]	58.102 (5.34) [0.116]	57.483 (6.56) [0.268]
Hollingshead	0.631 (1.13) [0.369]	0.233 (1.14) [0.331]	0.435 (1.08) [0.425]	0.322 (1.15) [0.419]
Group (1 = Mild/moderate injury)	−0.286 (0.10) [0.040]	−0.229 (0.10) [0.059]	−0.262 (0.09) [0.036]	−0.270 (0.10) [0.047]
Sex (1 = male)	−1.619 (1.71) [0.046]	−1.551 (1.74) [0.049]	−1.420 (1.72) [0.037]	−1.481 (1.81) [0.060]
Age	0.416 (1.74) [0.137]	0.643 (1.74) [0.166]	0.468 (1.73) [0.215]	0.247 (1.88) [0.182]
Race (1 = Caucasian)	0.361 (0.37) [0.133]	0.245 (0.41) [0.113]	0.245 (0.40) [0.176]	0.262 (0.45) [0.301]
Maternal Education	−0.580 (2.16) [0.255]	−1.387 (2.15) [0.230]	−0.868 (2.11) [0.303]	−0.321 (2.47) [0.258]
Error variance	109.520 (11.86)	112.019 (12.15)	110.357 (11.85)	121.777 (28.96)
Random effect variance	3.998 (2.91)	8.995 (3.56)	8.223 (10.04)	20.917 (53.39)