STATISTICS IN MEDICINE
Statist. Med. 2001; **20**:1755–1769 (DOI: 10.1002/sim.816)

A multiple imputation method for missing covariates in non-linear mixed-effects models with application to HIV dynamics

Hulin Wu[1,*,†] and Lang Wu[2]

[1] *Statistical and Data Analysis Center, Harvard School of Public Health, Frontier Science & Technology Research Foundation, Inc., 1244 Boylston Street, Suite 303, Chestnut Hill, Massachusetts 02467, U.S.A.*
[2] *Department of Biostatistics, Harvard School of Public Health, 655 Huntington Avenue, Boston, Massachusetts 02115-6023, U.S.A.*

SUMMARY

We propose a three-step multiple imputation method, implemented by Gibbs sampler, for estimating parameters in non-linear mixed-effects models with missing covariates. Estimates obtained by the proposed multiple imputation method are compared to those obtained by the mean-value imputation method and the complete-case method through simulations. We find that the proposed multiple imputation method offers smaller biases and smaller mean-squared errors for the estimates of covariate coefficients compared to other two methods. We apply the three missing data methods to modelling HIV viral dynamics from an AIDS clinical trial. We believe that the results from the proposed multiple imputation method are more reliable than that from the other two commonly used methods. Copyright © 2001 John Wiley & Sons, Ltd.

1. INTRODUCTION

For repeated measurement data, non-linear modelling is often required for meaningful analysis in many biological applications such as pharmacokinetic analysis and studies of growth and decay. The intra-individual variation and the inter-individual variation are usually modelled by a two-stage hierarchical model. The first stage specifies the mean and covariance structure for a given individual, while the second stage characterizes the inter-individual variation. Such models are often referred to as non-linear mixed-effects models or hierarchical non-linear models [1]. Understanding the nature of inter-individual systematic and random variation at the second stage often receives far more emphasis. Much of this inter-individual variation may be explained by covariates such as demographic and physiological information. Wu *et al.* [2] and Wu and Ding [3] proposed non-linear mixed-effects models for modelling HIV viral

*Correspondence to: Hulin Wu, Frontier Science & Technology Research Foundation, Inc., 1244 Boylston Street, Suite 303, Chestnut Hill, MA 02467-2115, U.S.A.
†E-mail: wu@sdac.harvard.edu

Contract/grant sponsor: NIAID/NIH; contract/grant number: AI 43220, AI 45356, AI38855

Received October 1999
Accepted July 2000

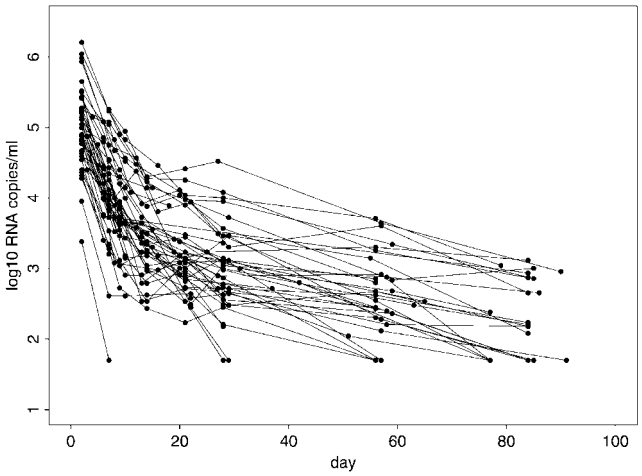

Figure 1. Plasma HIV-1 RNA copies ($\log_{10}$ scale) for ACTG 315 data. The solid circles are the observed values.

dynamics. These models have resulted in better understanding of the intra- and inter-individual variations in HIV dynamics and are important for further development of viral dynamic studies.

Figure 1 shows the profiles of HIV viral load measurements (plasma HIV-1 RNA copies) for 48 evaluable patients enrolled in the AIDS Clinical Trials Group (ACTG) Protocol 315. In this study, 53 HIV-1 infected patients are treated with potent antiviral treatment which consists of protease inhibitor (PI) and reverse transcriptase inhibitor (RTI) drugs (ritonavir monotherapy for the first 10 days, 3TC and AZT added on day 10). Five patients discontinued the study due to drug intolerance and other problems. The plasma HIV-1 RNA (viral load) is repeatedly quantified on days 0, 2, 7, 10, 14, 21, 28, and weeks 8, 12, 24 and 48 after initiation of treatment. We only consider the viral load data before rebound (presumably due to drug resistance and non-compliance). The data after 3 months are also excluded since these data are likely to be contaminated by the long-term clinical factors and may not be easily modelled using a parametric function. More detailed descriptions of the study can be found in Lederman *et al.* [4], Wu *et al.* [5] and Wu and Ding [3].

As is evident from Figure 1, the inter-patient variations appear to be large. Understanding the mechanism of the inter-patient variation is important for clinical decisions and treatment individualization. Preliminary study suggests that the inter-patient variation may be partially explained by some baseline (host) factors such as CD4 cell counts, natural killer activity (measured by lytic units or LU20), tumour necrosis factor (measured by plasma TNF levels), total complement levels (measured by CH50), and other covariates [5]. Many of these covariates, however, contain missing data. For example, the baseline values of LU20, CH50 and TNF contain 37.5 per cent, 19.0 per cent, and 16.7 per cent missing data in the study of ACTG 315, respectively. The main reason is that the study is a multi-centre study and some of the covariates are not measured at some centres. Thus, the missingness may not depend on the covariate values being missing, that is, the missing pattern may be assumed to be *missing at random* (MAR) in the sense of Rubin [6].

For non-linear mixed-effects models with missing covariates, the commonly used *complete-case method*, which discards all incomplete observations, is known to be inefficient. For example, for the ACTG 315 data, the complete-case method would discard more than 40 per cent of observations. Moreover, the complete-case method may even produce biased results if the subjects with complete data are not a random subsample of the entire sample (that is, if the missing data are not missing completely at random in the sense of Rubin [6]). Another common approach to handle missing data is to impute the missing values with the sample means of the partially observed covariates. This method, called the *mean-value imputation method*, may seriously dampen relationships among the covariates so may produce misleading covariate effects (see simulation results in Section 3).

Multiple imputation methods [12] take the uncertainty of the missing data into account, so often produce more reliable inference than single imputation methods. Although in multiple imputations we may need to assume a joint distribution for the partially observed variables, the method is often quite *robust* against moderate misspecification of the assumed model [7, 8]. One reason may be that failure of an imputation model does not damage the integrity of the entire data set, but only the portion that is imputed. Furthermore, as pointed out by Little [9], model-based methods for missing data problems are generally better than non-model-based methods even though model-based methods make distributional assumptions. Despite its many advantages, the multiple imputation method has not been widely used because of the lack of computational tools for generating multiple imputations. Recently, however, with the development of Markov chain Monte Carlo (MCMC) methods, implementation of multiple imputations in general missing data problems becomes possible.

Little [9] gives an excellent survey on missing covariates in regressions. The literature has primarily focused on missing covariates in linear models. Xie and Paik [8] studied multiple imputation methods for missing covariates in generalized linear models. Multiple imputation methods for missing data in the responses of a hierarchical survey model are proposed in Gelman *et al.* [10]. A related Bayesian approach to the problem of missing covariates in linear hierarchical models has been given by Dominici *et al.* [11]. These multiple imputation methods, as well as the standard imputation methods described in Rubin [12] and Schaffer [7], however, are not applicable to the missing covariates problem in non-linear mixed-effects models (or non-linear hierarchical models) because of either the non-linear relationships or the hierarchical structures. In this paper, we propose a multiple imputation method for missing covariates in non-linear mixed-effects models. The proposed method imputes the missing data at the individual level but can pool information across individuals.

In Section 2 we propose a multiple imputation method for missing covariates in non-linear mixed-effects models. Section 3 presents simulation results for comparing the proposed multiple imputation method with the mean-value imputation method and the complete-case method. In Section 4, we analyse the ACTG 315 data using the three missing data methods. The conclusion is given in Section 5.

2. A MULTIPLE IMPUTATION METHOD

2.1. The missing covariate problem

We consider the following non-linear mixed-effects model. Suppose we have an independent sample of n individuals, with individual i having p_i measurements. For individual i, let

$\mathbf{y}_i = (y_{i1}, \ldots, y_{ip_i})'$ be a $(p_i \times 1)$ vector of responses and $\mathbf{e}_i = (e_{i1}, \ldots, e_{ip_i})'$ be a $(p_i \times 1)$ vector of random intra-individual errors. Let the $(v \times 1)$ vector $\mathbf{t}_{ij}$ incorporate independent variables such as time, dose, etc. for individual i at measurement j. Then a non-linear mixed-effects model can be written as:

Stage 1: *intra-individual variation*

$$y_{ij} = f(\mathbf{t}_{ij}, \beta_i) + e_{ij}, \quad E(e_{ij}|\beta_i) = 0, \quad j = 1, \ldots, p_i, \quad i = 1, \ldots, n \tag{1}$$

with a specification of a covariance structure for e_{ij} which may depend on f and other parameters;

Stage 2: *inter-individual variation*

$$\beta_i = A_i\beta + \mathbf{u}_i, \quad \mathbf{u}_i \text{ i.i.d. } \sim \mathrm{N}(\mathbf{0}, D) \tag{2}$$

where $\beta_i = (\beta_{i1}, \ldots, \beta_{is})'$ is a $(s \times 1)$ vector representing the values of the regression parameters for individual i, A_i is a $(s \times r)$ design matrix whose elements are 0, 1, and $\mathbf{z}_i$, $\mathbf{z}_i = (z_{i1}, \ldots, z_{iq_i})'$ is a $(q_i \times 1)$ time-independent covariate vector corresponding to individual attributes for individual i, $\beta = (\beta_1, \ldots, \beta_r)'$ is a $(r \times 1)$ vector of fixed parameters (or fixed effects), and $\mathbf{u}_i = (u_{i1}, \ldots, u_{is})'$ is a $(s \times 1)$ vector of random effects associated with the ith individual. The covariance matrix D quantifies the random inter-individual variation. Note that the design matrix A_i which allows the same set of covariates for each component of the parameter vector β_i can be written as $A_i = I_s \otimes \mathbf{z}_i'$.

More general forms of hierarchical non-linear models can be found in Davidian and Giltinan [1]. For example, a more general form of the inter-individual variation at stage 2 is

$$\beta_i = \mathbf{d}(\mathbf{z}_i, \beta, \mathbf{u}_i), \quad \mathbf{u}_i \text{ i.i.d. } \sim (\mathbf{0}, D) \tag{3}$$

where $\mathbf{d}$ is an s-dimensional vector-valued function. The linear form (2) is most popular in practice and is our focus in this paper. Our method, however, is still applicable to models with more general form (3). In some cases, a non-linear form of (3) can be converted into the linear form of (2) by variable transformations.

We consider the missing covariate problem for the non-linear mixed-effects models (1) and (2). We assume that all y_{ij} are observed. The missing data are assumed to be missing at random in the sense of Rubin [6], which is reasonable for the ACTG 315 data as noted in Section 1. We also assume that $q_1 = q_2 = \cdots = q_n = q$, that is, the same set of covariates is used to explain the inter-individual variation for each individual. Otherwise we can treat the absence of certain covariates for a given individual as missing data.

2.2. The imputation model

Suppose a total of $q + q^*$ covariates are measured for each of the n individuals. For the first q covariates, missing values are present on the n observations for each covariate (that is, these q covariates are incompletely observed), while for the remaining q^* covariates, no missing values are present on the n observations for any covariate (that is, these q^* covariates are completely observed). Let Z be the $n \times q$ *incompletely* observed covariate matrix whose ith row is $\mathbf{z}_i = (z_{i1}, \ldots, z_{iq})$, and let X denote the $n \times q^*$ *completely* observed covariates matrix whose ith row is $\mathbf{x}_i = (x_{i1}, \ldots, x_{iq^*})$. A simple multiple imputation (MI) method for missing covariates would be to create imputations based on an assumed joint distribution (usually normal distribution) for the covariates (Z, X) or based on an assumed

conditional distribution for the incompletely observed covariates given the completely observed covariates $(Z|X)$ (so to avoid distributional assumption for the completely observed covariates X and reduce the number of parameters). A drawback of this approach is that it ignores the information contained in the response variable $\mathbf{y}$ regarding the incompletely observed covariates Z. For non-linear models, a joint normal distribution assumption for the covariates (Z,X) and the response $\mathbf{y}$ may be also unrealistic because of the non-linear relationship between the covariates and the response. As suggested by Little [9] for missing covariates in general regression problems and considered in Xie and Paik [8], a more appropriate MI model is to create imputations based on an assumed conditional distribution for $(Z|X,\mathbf{y})$. By conditioning on $(X,\mathbf{y})$, we avoid distributional assumption in the imputations not only for the completely observed covariates X, but also for the response $\mathbf{y}$. The idea of MI methods is then to generate random samples from the assumed predictive distribution of $(Z|X,\mathbf{y})$ to impute the missing values in Z.

For a simple missing covariate problem in generalized linear models with special missing data patterns, Xie and Paik [8] have proposed to generate multiple imputations from the linear model $(Z|X,\mathbf{y})=(X,\mathbf{y})\times\eta+\varepsilon$, where η is an unknown parameter vector and ε is normally distributed. This imputation model cannot be easily extended to non-linear mixed-effects models since the numbers of repeated measurements of the response are different across individuals. Moreover, for non-linear regressions the relation between the response and the covariates is often complicated, as is obvious from (1) and (2), so the linear imputation model assumed in Xie and Paik [8] may not be adequate. For the non-linear mixed-effects models (1) and (2), we propose an MI procedure as follows. First we fit the non-linear mixed-effects model without any covariates in the second stage (2), and obtain initial individual estimates $\hat{\beta}_i^{(0)}$ of β_i. These initial individual estimates $\hat{\beta}_i^{(0)}$ contain the information in the response y_{ij} about the covariates [1, 13], and are sufficient statistics for β. Treating $\hat{\beta}_i^{(0)}$ as observed data of β_i, which is a common approach [1], the linear form of (2) $\beta_i=A_i\beta+\mathbf{u}_i$ suggests that the relation between β_i and the covariates $\mathbf{z}_i$ is linear, so it is appropriate to assume a linear imputation model for $(Z|X,\hat{\beta}_i^{(0)})$. Specifically, we propose the following three-step procedure to generate multiple imputations for the partially observed covariates Z, and to estimate the parameters in the non-linear mixed-effects model:

Step 1. Fit the hierarchical model without covariates

$$y_{ij}=f(\mathbf{t}_{ij},\beta_i)+e_{ij}$$
$$\beta_i=\beta_0+\mathbf{u}_i,\quad i=1,\ldots,n;\ j=1,\ldots,p_i$$

Denote the resulting individual estimates of β_i by $\hat{\beta}_i^{(0)}=(\hat{\beta}_{i1}^{(0)},\ldots,\hat{\beta}_{is}^{(0)})'$.

Step 2. Impute the missing values in Z based on the multivariate linear model

$$Z=(\mathbf{1}\ X\ \hat{\beta}^{(0)})B+\varepsilon=WB+\varepsilon \tag{4}$$

where $\mathbf{1}=(1,\ldots,1)'$, $W=(\mathbf{1}\ X\ \hat{\beta}^{(0)})$ is an $n\times(1+q^*+s)$ matrix whose ith row is $(1,x_{i1}\cdots x_{iq^*}\hat{\beta}_{i1}^{(0)}\cdots\hat{\beta}_{is}^{(0)})$, $\hat{\beta}^{(0)}$ is an $n\times s$ matrix whose ith row is $(\hat{\beta}_{i1}^{(0)}\cdots\hat{\beta}_{is}^{(0)})$, $B=(b_{ij})$ is a $(1+q^*+s)\times q$ matrix of parameters, and ε is an $n\times q$ random matrix whose n rows are independently distributed as $\mathrm{N}_q(\mathbf{0},\Sigma)$. Let $Z=(Z_{\mathrm{obs}},Z_{\mathrm{mis}})$, where Z_{obs} and Z_{mis} denote the

observed values and the missing values in Z, respectively. We generate multiple imputations by drawing m *independent* samples, $Z_{\text{mis}}^{(1)}, \ldots, Z_{\text{mis}}^{(m)}$, from $P(Z_{\text{mis}} \mid Z_{\text{obs}}, W)$ according to the above model (see next subsection for the details of implementation).

Step 3. After creating m independent complete data sets $\{(Z_{\text{obs}}, Z_{\text{mis}}^{(j)}),\ j = 1, \ldots, m\}$, we can analyse each data set by standard complete data methods. The m analysis results are then combined to obtain an overall inference. For example, if $\tilde{\beta}^{(j)}$ and $\tilde{V}^{(j)}$ are, respectively the estimate of β and its associated variance based on the complete data $(Z_{\text{obs}}, Z_{\text{mis}}^{(j)})$, the overall estimates of β and its associated variance are, respectively

$$\tilde{\beta} = \frac{1}{m}\sum_{j=1}^{m} \tilde{\beta}^{(j)}, \quad \tilde{V} = \frac{1}{m}\sum_{j=1}^{m} \tilde{V}^{(j)} + \frac{m+1}{m(m-1)}\sum_{j=1}^{m} (\tilde{\beta}^{(j)} - \tilde{\beta})^2$$

See Rubin [12] and Schafer [7] for details on combining estimates, standard errors, p-values, and other related inferences.

After creating multiple imputations in steps 2 and 3, we may go back to step 1, with the second stage in the hierarchical model replaced by (2) and the missing covariates replaced by the imputed values, and obtain an updated estimate of β_i (average over the m estimates), denoted by $\hat{\beta}_i^{(1)}$. Then in step 2, we replace $\hat{\beta}_i^{(0)}$ in W by $\hat{\beta}_i^{(1)}$ and create updated imputations. We may iterate these steps until convergence is achieved in $\tilde{\beta}$. This iterative approach, however, may substantially increase computational burden. An alternative approach to address the uncertainty in estimating $\hat{\beta}^{(0)}$ is to assume a joint distribution for $(Z, \hat{\beta}^{(0)})$, such as a multivariate normal distribution, and then create multiple imputations for the missing values in Z in a standard manner. However, this approach may substantially increase the number of unknown parameters in the imputation model and require distributional assumption for $\hat{\beta}^{(0)}$. Our preliminary simulations show that these two procedures provide little improvements, presumably because multiple imputation procedures are generally robust against small to moderate misspecification of the imputation models and these procedures do not seem to substantially improve the imputation model. Thus, the imputation model (4) can be used to create reasonable imputations. Iterations and further refinements may not be necessary if they increase the computation time or the number of unknown parameters substantially.

In the imputation model (4), we considered the initial parameter estimates $\hat{\beta}_i^{(0)}$ from step 1 and the completely observed covariates X that may contain information (association) regarding missing values in Z. The association among the partially observed covariates Z is naturally incorporated via the unstructured covariance matrix Σ. These associations are used to predict the missing values in Z. By creating m imputations for each missing value, we take the uncertainty of the prediction into account. Because the m imputations are generated independently, the multiple imputation is (Bayesianly) proper (Schaffer [7]), so the resulting inference is valid.

We allow an arbitrary number of covariates and arbitrary missing data patterns. By regressing on the completely observed covariates in modelling partially observed covariates, the model allows the completely observed covariates to be categorical, continuous, or a mix of both types, and allows interactions or polynomials or other non-linear functions. For incompletely observed covariates, although we assume a conditional multivariate normal distribution, our method is still applicable when some covariates are categorical. When some incompletely observed covariates are categorical (for example, binary or ordinal), the simplest approach

is to model the data as continuous and impute continuous values, then we round off the continuous imputes to the nearest category. This approach is often reasonable [7, 10].

Note that, although MI methods are generally robust against moderate misspecification of the imputation models, it is still important to check the adequacy of the imputation models in order to improve the prediction of the missing values. A useful approach would be to use cross-validation methods to check the adequacy of the imputed values, as is demonstrated in Gelman *et al.* [10]. In Section 3 we will present some simulations to compare the proposed MI method with other commonly used methods.

2.3. Implementation using Gibbs sampler

The implementation of the multiple imputation is described as follows. We generate multiple imputations by drawing m independent samples, $Z_{\text{mis}}^{(1)}, \ldots, Z_{\text{mis}}^{(m)}$, from $P(Z_{\text{mis}}|Z_{\text{obs}}, W)$ according to (4). Note that the posterior distribution of the missing data can be written as

$$P(Z_{\text{mis}}|Z_{\text{obs}}, W) = \int P(Z_{\text{mis}}|\theta, Z_{\text{obs}}, W)P(\theta|Z_{\text{obs}}, W)\,\mathrm{d}\theta \tag{5}$$

where $\theta = (B, \Sigma)$, $P(\theta|Z_{\text{obs}}, W) = \int L(\theta|Z_{\text{obs}}, Z_{\text{mis}}, W)\,\mathrm{d}Z_{\text{mis}}$, and $L(\theta|Z_{\text{obs}}, Z_{\text{mis}}, W)$ is the likelihood function for complete data. Thus we must specify a prior distribution for θ. After imputations, we create m independent complete data sets $\{(Z_{\text{obs}}, Z_{\text{mis}}^{(j)}, X),\ j = 1, \ldots, m\}$, each of which can be used in the fitting of the non-linear mixed-effects model by standard complete data methods. The results of the m analyses are then combined to obtain inferences that efficiently incorporate the uncertainty about Z_{mis}.

A common non-informative prior for $\theta = (B, \Sigma)$ is

$$P(B) \propto \text{constant}, \quad P(\Sigma) \propto |\Sigma|^{-\frac{q+1}{2}} \tag{6}$$

However, the improper prior for the covariance matrix Σ in (6) may lead to Gibbs samplers that do not converge to proper posteriors. Thus we may consider the following proper prior distribution for the covariance matrix Σ:

$$\Sigma \sim W^{-1}(\gamma, \Lambda), \quad \gamma \geqslant q \tag{7}$$

where $W^{-1}(\cdot)$ denotes the inverted-Wishart distribution. In choosing the hyperparameters, we may think $\gamma^{-1}\Lambda^{-1}$ as a prior guess for Σ. Note that a small value for γ can reduce its impact on the final inferences. If we treat Z as completely observed, we can apply the Bayes theorem to obtain the following posterior distributions:

$$\Sigma \mid Z, W \sim \mathscr{W}^{-1}(n - s - q - q^*, \hat{\varepsilon}'\hat{\varepsilon}), \tag{8}$$

$$B \mid Z, W, \Sigma \sim N(\hat{B}, (W'W)^{-1} \otimes \Sigma) \tag{9}$$

$$Z \mid W, B, \Sigma \sim N(WB, I_n \otimes \Sigma) \tag{10}$$

where

$$\hat{\varepsilon} = Z - W\hat{B} \tag{11}$$

$$\hat{B} = (W'W)^{-1}(W'Z) \tag{12}$$

We use the Gibbs sampler [14–16] to generate the predictive values of Z_{mis}. Note that, because the rows $\mathbf{z}_1, \ldots, \mathbf{z}_n$ of Z are independent given (B, Σ), we can write

$$P(Z_{\text{mis}}|Z_{\text{obs}}, W, B, \Sigma) = \prod_{i=1}^{n} P(\mathbf{z}_{i(\text{mis})}|\mathbf{z}_{i(\text{obs})}, W, B, \Sigma)$$

where $\mathbf{z}_{i(\text{mis})}$ and $\mathbf{z}_{i(\text{obs})}$ denote the missing and observed subvectors of $\mathbf{z}_i$, respectively. The distribution $P(\mathbf{z}_{i(\text{mis})}|\mathbf{z}_{i(\text{obs})}, W, B, \Sigma)$ is multivariate normal which is equivalent to a linear regression model, regressing $\mathbf{z}_{i(\text{mis})}$ on $\mathbf{z}_{i(\text{obs})}$, and the parameters of this regression can be calculated by the sweeping operator described in Little and Rubin [17]. Given a current estimate of $(B^{(t)}, \Sigma^{(t)})$, each iteration of the Gibbs sampler consists of the following steps:

Step 1. Given $(B^{(t)}, \Sigma^{(t)})$ and $\mathbf{z}_{i(\text{obs})}$, draw a value of $\mathbf{z}_{i(\text{mis})}^{(t+1)}$ from $P(\mathbf{z}_{i(\text{mis})}|\mathbf{z}_{i(\text{obs})}, W, B^{(t)}, \Sigma^{(t)})$ independently for $i = 1, 2, \ldots, n$, thus we obtain $Z_{\text{mis}}^{(t+1)}$.

Step 2. Given $Z_{\text{obs}}, Z_{\text{mis}}^{(t+1)}$ and W, draw a value of $\Sigma^{(t+1)}$ from $P(\Sigma|Z_{\text{obs}}, Z_{\text{mis}}^{(t+1)}, W)$ based on (8).

Step 3. Given $Z_{\text{obs}}, Z_{\text{mis}}^{(t+1)}, W$, and $\Sigma^{(t+1)}$, draw a value of $B^{(t+1)}$ from $P(B|Z_{\text{obs}}, Z_{\text{mis}}^{(t+1)}, W, \Sigma)$ based on (9).

Repeating these three steps from a starting value $(B^{(0)}, \Sigma^{(0)})$ yields a stochastic sequence $\{Z_{\text{mis}}^{(t)}: t = 1, 2, \ldots\}$, which has $P(Z_{\text{mis}}|Z_{\text{obs}}, W)$ as its stationary distribution. Thus, for a sufficiently large value of t, we can regard $Z_{\text{mis}}^{(t)}$ as an approximate draw from $P(Z_{\text{mis}}|Z_{\text{obs}}, W)$. Note that values from inverted-Wishart distribution $\mathcal{W}_p^{-1}(n, \Lambda)$ can be generated by noting that $\mathcal{W}_p^{-1}(n, \Lambda) = (V'V)^{-1}$, where V is an $n \times p$ random matrix whose rows are independently distributed as $\text{N}(0, \Lambda)$. Convergence of the MCMC methods can be informally assessed by examining the time-series plots, autocorrelations etc., see Schafer [7] and Gilks *et al*. [18] for discussion on formal and informal convergence diagnostics for MCMC.

3. SIMULATIONS

In this section we use simulations to compare three methods for missing covariates in non-linear hierarchical models – the complete-case method, the mean-value imputation method and the multiple imputation method proposed in this paper. The covariates in the simulations are assumed to be time-independent. We conduct our simulation studies using a non-linear mixed-effects model of HIV-1 dynamics proposed by Wu and Ding [3]. The design of this simulation experiment and the true values of the parameters assumed in the simulation are based on the real data (ACTG 315) example which is introduced in Section 1 and will be analysed in the next section.

We generate data from the following non-linear hierarchical model with two covariates (Wu and Ding [3]):

$$\begin{aligned}
y_{ij} &= \log_{10}(P_{1i}\text{e}^{-\lambda_{1i}t_{ij}} + P_{2i}\text{e}^{-\lambda_{2i}t_{ij}}) + e_{ij} \\
P_{1i} &= P_{10} + a_{11}z_{1i} + a_{12}z_{2i} + u_{1i} \\
\lambda_{1i} &= \lambda_{10} + a_{21}z_{1i} + a_{22}z_{2i} + u_{2i} \\
P_{2i} &= P_{20} + a_{31}z_{1i} + a_{32}z_{2i} + u_{3i} \\
\lambda_{2i} &= \lambda_{20} + a_{41}z_{1i} + a_{42}z_{2i} + u_{4i}, \quad j = 1, \ldots, p; \; i = 1, \ldots, n
\end{aligned} \tag{13}$$

Thus $\beta_i = (P_{1i}, \lambda_{1i}, P_{2i}, \lambda_{2i})$ and $\beta = (P_{10}, a_{11}, a_{12}, \lambda_{10}, a_{21}, a_{22}, P_{20}, a_{31}, a_{32}, \lambda_{20}, a_{41}, a_{42})$ are the individual-specific parameters and the population parameters, respectively. We assume that $e_{ij}|\beta_i \sim \mathrm{N}(0, \Lambda)$ and $\mathbf{u}_i = (u_{1i}, u_{2i}, u_{3i}, u_{4i}) \sim \mathrm{N}(0, D)$. The number of individuals is $n = 50$, and the number of measurements for each individual is $p = 8$ ($t = 2, 3, 4, 6, 8, 12, 16, 25$). The true values of the parameters are $\beta = (12.5, 1.0, 2.0, 0.8, 0.1, 0.1, 8.0, 1.0, 1.0, 0.2, 0.03, 0.03)$. We assume a simple diagonal matrix for the within-patient covariance matrix Λ, with a constant variance $\sigma^2 = 0.265$. The true value of the matrix D is

$$D = \begin{pmatrix} 1.21 & -0.02 & 1.48 & 0.01 \\ -0.02 & 0.015 & -0.005 & 0.001 \\ 1.48 & -0.005 & 2.25 & 0.01 \\ 0.01 & 0.001 & 0.01 & 0.0003 \end{pmatrix}.$$

We consider two missing data mechanisms in our simulation studies: missing completely at random (MCAR) and missing at random (MAR). For the case of MCAR, for each simulated sample we *randomly* delete some values of z_1 and z_2 such that the two covariates z_1 and z_2 have the same missing rates. For the case of MAR, for each simulated sample we delete those values (with probability 0.9) of z_1 which correspond to the largest values of the response y_{ij}, and we delete those values (with probability 0.9) of z_2 which correspond to the smallest values of the response y_{ij}, since in our real data set some covariates corresponding to large or small responses appear more likely to be missing. The missing rates for z_1 and z_2 are again the same. We consider two missing rates for z_1 and z_2 in the study: 20 per cent missing and 40 per cent missing; 500 replicates are run.

The number of imputations for the MI method is $m = 6$ (See Schafer [7] for reasonable numbers of imputations for the MI methods; often $m \leqslant 6$ may be sufficient.) The starting values for the hyperparameters of the prior distribution of the covariance matrix Σ are chosen to be $\gamma = q$ and $\Lambda = I$. We found that this informative prior and the non-informative prior almost produce the same inference. To determine the burn-in or warm-up iterations in the Gibbs sampler, we run some preliminary simulations first. We assess the convergences by examining time-series plots and autocorrelations. We found that the chains converged rapidly; all achieved stationarity within 100 iterations. To ensure the convergence safely, we conservatively chose 500 iterations for burn-in for each MCMC sample.

The estimates from four methods are compared: the estimates obtained using full data (FD); the estimates obtained by the complete-case method (CC); the estimates obtained by the mean-value imputation method (MV), and the estimates obtained by the multiple imputation method proposed in this paper (MI). We use the S-plus function NLME, which is a method based on linearization and is developed by Pinheiro and Bates [19], to obtain the MLEs of the parameters. We compute the bias and mean-squared error (MSE) for each estimate. Since we mainly focus on the biases of the estimates from different methods and since the standard errors of the estimates obtained from the linearization method used to fit the models are incorrect [1], we do not report the standard errors in our simulation studies. The bias for the jth component of β is defined as $\mathrm{bias}_j = \hat{\beta}_j - \beta_j$. The mean-squared error for the jth component of β is defined as $\mathrm{MSE}_j = \mathrm{bias}_j^2 + s_j^2$, where s_j is the simulated standard error of $\hat{\beta}_j$. To better compare the results, we calculate per cent bias, defined as $100 \times \mathrm{bias}_j/\beta_j$, and percent MSE,

Table I. Per cent biases and mean-squared errors of the four estimation methods when the missing data are MCAR.

Missing rate	Parameters (true values)	Per cent bias				Per cent MSE			
		FD	MI	MV	CC	FD	MI	MV	CC
20%	$a_{11} = 1.0$	−4	2	41*	−7	9	8	42	15
	$a_{12} = 2.0$	1	−4	−4	2	4	5	6	6
	$a_{21} = 0.1$	−4	−8	13	−16	18	17	23	36
	$a_{22} = 0.1$	4	2	5	2	17	16	20	34
	$a_{31} = 1.0$	1	−3	16*	−4	14	13	21	23
	$a_{32} = 1.0$	4	9	14	15	15	16	19	28
	$a_{41} = 0.03$	2	−2	18*	4	18	17	26	37
	$a_{42} = 0.03$	5	10	17	21	18	21	24	42
40%	$a_{11} = 1.0$	−1	3	99*	−2	8	13	100	27
	$a_{12} = 2.0$	−1	−4	13	−6	4	12	14	15
	$a_{21} = 0.1$	−4	−9	41*	−9	20	17	44	59
	$a_{22} = 0.1$	1	−4	37*	−17	20	18	41	60
	$a_{31} = 1.0$	−1	−3	47*	−15	13	13	49	45
	$a_{32} = 1.0$	8	6	47*	27	15	14	49	50
	$a_{41} = 0.03$	−1	−1	45*	−9	17	17	49	61
	$a_{42} = 0.03$	10	10	50*	28	20	19	53	68

*Biases significantly different from zero.

defined as $100 \times \sqrt{(\text{MSE})_j}/\beta_j$. We focus on how different missing data methods affect the estimation of the covariate effects a_{ij}.

Table I summarizes the simulation results for the MCAR missing pattern. The values in the tables are averages from the 500 simulations. For the three missing covariate methods (MI, MV and CC), Table I shows that the estimates from the proposed MI method have the smallest MSEs consistently for all cases compared to the MV and CC methods. The biases of the MI estimates are also smaller for most parameters compared to the MV and CC methods. Table I also indicates that the MV method is the worst in the sense that it may produce severely biased estimates and may give misleading covariate effects. The CC method offers relatively small biases in the MCAR case, but the MSEs of the CC estimates are substantially large, especially when the missing rate is high. This is expected since the CC method discards all partially observed data.

Simulation results for the MAR missing pattern are given in Table II. In this case, the missingness of the covariates depends on the values of the responses. Table II shows that now both MV and CC methods give severely biased estimates, so both methods would produce misleading covariate effect estimates and incorrect individual-specific parameter estimates. The proposed MI method, on the other hand, still offers the smallest biases and MSEs.

Note that, due to random variation and convergence problems, the FD and MI methods may show somewhat large biases in a few cases compared to the MV and CC methods, but these biases are not significant. The overall performance of the estimates should be better assessed by the MSEs. In terms of the MSE, the MI method is consistently superior to the MV and CC methods for all parameters and in all our simulation cases, and is highly efficient relative to the FD estimates.

Table II. Per cent biases and mean-squared errors of the four estimation methods when the missing data are MAR.

Missing rate	Parameters (true values)	Per cent bias				Per cent MSE			
		FD	MI	MV	CC	FD	MI	MV	CC
20%	$a_{11} = 1.0$	−1	6	77*	−34*	8	10	78	35
	$a_{12} = 2.0$	−1	−5	5	−35*	4	6	8	36
	$a_{21} = 0.1$	−4	−11	46*	−23	18	18	48	38
	$a_{22} = 0.1$	1	4	32*	−20	19	15	37	32
	$a_{31} = 1.0$	6	6	40*	−39*	14	12	41	42
	$a_{32} = 1.0$	5	8	12	−55*	14	14	24	59
	$a_{41} = 0.03$	10	6	58*	−17	19	18	59	32
	$a_{42} = 0.03$	6	12	27	23	18	20	40	33
40%	$a_{11} = 1.0$	2	11	107*	−81*	8	15	108	83
	$a_{12} = 2.0$	−1	−13	21*	−77*	4	14	22	78
	$a_{21} = 0.1$	−2	−15	73*	−43*	21	21	76	81
	$a_{22} = 0.1$	−2	2	74*	−32*	20	16	76	79
	$a_{31} = 1.0$	5	−4	51*	−70*	16	14	53	81
	$a_{32} = 1.0$	3	1	29*	−138*	14	15	35	145
	$a_{41} = 0.03$	3	−2	83*	−70*	19	19	86	72
	$a_{42} = 0.03$	3	12	67*	−32	17	22	73	82

*Biases significantly different from zero.

4. APPLICATION TO HIV-1 DYNAMICS

In this section we analyse the ACTG 315 data described in Section 1. Wu and Ding [3] recommended modelling the data by the biphasic model (13), where y_{ij} is the $\log_{10}$ transformation of the HIV viral load measurements for the ith patient at jth time point. The log-transformation of the raw viral load data is used to stabilize the variance and to make the raw data more normally distributed. The two viral decay rates λ_{1i} and λ_{2i} represent the turnover rates of productively infected cells and long-lived and/or latently infected cells, respectively. P_{1i} and P_{2i} are the amounts of virus produced and cleared from the corresponding cells, respectively. As in Wu and Ding [3], we assume that the variance-covariance matrix Λ is diagonal with constant variance. We also take log-transformations of the parameters P_{1i} and P_{2i} to ensure the positivity of these parameters, to make the components of β_i to have similar scales, and to make β_i more normally distributed.

As is indicated in Figure 1, the interpatient variation appears to be large. Much of the large interpatient differences may be explained by covariates such as demographic and physiological information (Wu *et al.* [5]). Based on HIV/AIDS clinicians' suggestions (Wu *et al.* [5]), patients with higher CD4 counts may have faster decay rates since more CD4 cells may help to eliminate infected cells; HIV-infected cells may be killed by killer cells, thus natural killer activity (measured by lytic units or LU20) may also affect the viral decay rates. Medical investigators also suggested that the tumour necrosis factor (measured by plasma TNF levels) are correlated with viral load, and that the total complement levels (as measured by CH50) increased during antiviral treatment. We consider these covariates in our HIV dynamic model in this exploratory analysis, although some other factors may also affect HIV dynamics. Since

Table III. Estimates of population parameters based on the three missing data methods for the ACTG 315 data.

Parameters	MI method		CC method		MV method	
	Estimate	SE	Estimate	SE	Estimate	SE
P_{10}	12.469	0.156	12.531	0.174	12.467	0.160
a_{11}	0.745	0.169	0.832	0.178	0.760	0.176
λ_{10}	0.481	0.024	0.464	0.026	0.475	0.024
a_{21}	0.043	0.031	0.073	0.033	−0.023	0.014
a_{22}	0.065	0.022	0.050	0.021	0.052	0.021
P_{20}	8.061	0.198	8.051	0.210	8.013	0.200
a_{31}	0.718	0.184	0.650	0.179	0.718	0.192
λ_{20}	0.041	0.003	0.041	0.003	0.040	0.003
a_{41}	0.006	0.003	0.005	0.002	0.006	0.002

the data are sparse for many individuals (the number of viral load measurements before viral rebound for each individual ranges from 2 to 8 measurements), we use the linearization procedure of Lindstrom and Bates [20] to fit various models using the S-plus function NLME. Standard model selection techniques based on complete data, such as the likelihood-ratio (LR) test, AIC or BIC criterion, lead to adoption of the following model for the interpatient variation (that is, the second stage of the non-linear mixed-effects model):

$$\log(P_{1i}) = \log(P_{10}) + a_{11}\mathrm{TNF} + u_{1i},$$
$$\lambda_{1i} = \lambda_{10} + a_{21}\mathrm{TNF} + a_{22}\mathrm{CD4} + u_{2i},$$
$$\log(P_{2i}) = \log(P_{20}) + a_{31}\mathrm{TNF} + u_{3i},$$
$$\lambda_{2i} = \lambda_{20} + a_{41}\mathrm{CH50} + u_{4i}.$$

We applied the three missing data methods (CC, MV and MI methods) to the ACTG 315 data. For the CC method, we discarded all incomplete observations. We also made a log transformation on the covariate LU20 to make its distribution more symmetric and closer to normality. Table III gives the estimates from the three methods. As noted earlier, the default standard errors from the output of S-plus function NLME (SE in Table III) may be inaccurate [1] and can only be used as a reference. Table III indicates that different missing data methods can lead to different estimation of a_{ij} – the covariate effects on the individual-specific parameters. Because of the large interpatient variations revealed in Figure 1, in clinical decisions we should treat each patient differently. Thus, it is important to accurately estimate the parameters for each individual. However, in the presence of missing data in covariates, individual parameter estimates can be quite different if different missing data methods are used. For instance, in this example, for a patient with baseline $\mathrm{CD4} = 127$, $\mathrm{CH50} = 270$ and $\mathrm{TNF} = 165$, the estimated first decay rates are $\lambda_{1i} = 0.60$ based on the MI method, $\lambda_{1i} = 0.70$ based on the CC method and $\lambda_{1i} = 0.36$ based on the MV method. For all patients in the study, the individual estimates of λ_{1i} range from 0.41 to 0.72 based on the MI method, from 0.32 to 0.87 based on the CC method and from 0.15 to 0.87 based on the MV method. The estimates of λ_{2i} range from 0.028 to 0.062 based on the MI method, from 0.023 to 0.070 based

Table IV. Estimates of population parameters based on the three missing data methods for the new data set.

Parameters	MI method		CC method		MV method	
	Estimate	SE	Estimate	SE	Estimate	SE
P_{10}	12.495	0.177	10.432	0.548	10.926	0.389
a_{11}	0.719	0.223	0.033	0.008	0.025	0.005
λ_{10}	0.487	0.030	0.217	0.099	0.159	0.080
a_{21}	0.056	0.038	0.002	0.001	0.001	0.001
a_{22}	0.081	0.033	0.0007	0.0003	0.001	0.0003
P_{20}	8.143	0.245	6.002	0.669	6.594	0.427
a_{31}	0.655	0.197	0.028	0.009	0.024	0.006
λ_{20}	0.043	0.004	0.012	0.012	0.010	0.014
a_{41}	0.005	0.003	0.0001	0.00005	0.0001	0.00006

on the CC method and from 0.017 to 0.049 based on the MV method. Thus, the selection of missing data methods is crucial for the estimates of covariate effects, in particular, for parameter estimates of the individuals with missing covariates.

From the results in Table III, the differences in population parameter estimates are not very large for different missing data methods, presumably due to the low missing rate of the included covariates (19 per cent in CH50, 17 per cent in TNF, 0 per cent in CD4). To further illustrate how the missing data methods perform differently when the missing rate is high, we deliberately deleted 40 per cent of the completely observed covariate CD4 by random and created a new data set. The foregoing analysis was repeated using the three missing data methods. The population estimates of parameters are given in Table IV. We can see that the parameter estimates from the CC and MV methods change dramatically from the previous analysis (Table III). The population parameter estimates from the MI method, however, are quite similar to the previous analysis. For the particular patient mentioned above (baseline CD4 = 127, CH50 = 270 and TNF = 165), the estimate of first-phase viral decay rate now become $\lambda_{1i} = 0.65, 0.22$ and 0.16 based on the MI, CC and MV methods, respectively. The MI estimate of λ_{1i} is only changed slightly, but the CC and MV estimates are changed substantially compared to the previous estimates. Thus, the differences in both population estimates and individual estimates are striking for the three different missing data methods.

As is demonstrated in the simulations in Section 3, the CC method is inefficient and may even be biased if the missing data are not missing completely at random. The MV method is even worse than the CC method. Simulation studies show that the proposed MI method is the most reliable method. To further check the adequacy of the MI model, we consider the following cross-validation method. First we create a new data set by randomly deleting 20 per cent of the observed CD4 values and 20 per cent of the observed TNF values. We then run the multiple imputation program on this new data set, and for each individual for whom we artificially removed the covariate values, we compare the five imputed values to the true value. If the data had been simulated from the model, then we would expect the actual values and the multiple imputations to have the same distribution, so that if one ranked the actual value along with five random imputations, all six possible orderings would be equally likely. Table V gives the cross-validation result. The result shows that the six rankings are

Table V. Cross-validation for the multiple imputation model.

Covariate	Rank of true value among the five imputations					
	1	2	3	4	5	6
CD4	18.4%	14.1%	17.4%	15.1%	20.2%	14.8%
TNF	16.4%	20.8%	20.1%	14.1%	13.5%	15.0%

close to equally likely. Thus the MI model should be appropriate, since, if the imputations were systematically too high or too low, the frequencies in the six categories would show a decreasing or increasing trend, respectively.

The above analyses based on the MI method suggest the need of individualized treatments for HIV-infected patients based on covariate information. In above analysis, we only considered three covariates. In fact, there are many other factors that may also affect the viral decay rates. It is also important to identify significant covariates among a large number of available covariates for prediction of viral decay rates. However, the covariate selection problem (with missing data) is out of the scope of this paper and is not addressed here. This will be one of our future research topics.

5. CONCLUSION

In non-linear mixed-effects models, when the inter-individual variation is large, it is desirable to consider covariates to explain these variations. When the covariates contain missing data, the commonly used complete-case method and the mean-value-imputation method may give biased and misleading results. We have proposed a multiple imputation method to handle the missing covariate problem in the non-linear mixed-effects models. Our simulation studies show that the proposed multiple imputation method is better than the other two methods in the sense that it offers smaller bias and smaller mean-squared errors for parameter estimates. The results based on the multiple imputation method appear to be more reliable.

Our proposed multiple imputation method is applicable to missing covariates problems with arbitrary missing data patterns and arbitrary number of covariates. Although our simulation studies and the real data example are based on an HIV viral dynamic model, the proposed method is generally applicable to other non-linear mixed-effects modelling areas such as pharmacokinetics and pharmacodynamics. We have focused on non-linear mixed-effects models with time-independent covariates. Methods for non-linear mixed-effects models with time-dependent missing covariates are currently under investigation.

ACKNOWLEDGEMENTS

This work was supported by NIAID/NIH grants AI43220, AI45356 and AI38855. We thank Dr Adam Ding and Professor Victor DeGruttola for helpful comments and suggestions. We also thank the two referees for their constructive and insightful comments and suggestions, which helped us to greatly improve the paper.

REFERENCES

1. Davidian M, Giltinan DM. *Nonlinear Models for Repeated Measurements Data.* Chapman & Hall, 1995.
2. Wu H, Ding A, De Gruttola V. Estimation of HIV dynamic parameters. *Statistics in Medicine* 1998; **17**: 2463–2485.
3. Wu H, Ding A. Population HIV-1 dynamics in vivo: applicable models and inferential tools for virological data from AIDS clinical trials. *Biometrics* 1999; **55**:410–418.
4. Lederman MM, Connick E, Landay A, Kuritzkes DR, Spritzler J, Clair MS, Kotzin BL, Fox L, Heath-Chiozzi M, Leonard JM, Rousseau F, Wade M, D'Arc Roe J, Martinez A, Kessler H. Immunologic responses associated with 12 weeks of combination antiretroviral therapy consisting of zidovudine, lamivudine and ritonavir: results of AIDS Clinical Trials Group Protocol 315. *Journal of Infectious Diseases* 1998; **178**:70–79.
5. Wu H, Kuritzkes DR, McClernon DR, Kessler H, Connick E, Landay A, Spear G, Heath-Chiozzi M, Rousseau F, Fox L, Spritzler J, Leonard JM, Lederman MM. Characterization of viral dynamics in human immunodeficiency virus type 1-infected patients treated with combination antiretroviral therapy: relationships to host factors, cellular restoration and virological endpoints. *Journal of Infectious Diseases* 1999; **179**(4):799–807.
6. Rubin DB. Inference and missing data. *Biometrika* 1976; **63**:581–592.
7. Schafer JL. *Analysis of Incomplete Multivariate Data.* Chapman & Hall, 1997.
8. Xie F, Paik MC. Multiple imputation methods for the missing covariates in generalized estimating equation. *Biometrics* 1997; **53**:1538–1546.
9. Little RJA. Regression with missing *X*'s: a review. *Journal of the American Statistical Association* 1992; **87**: 1227–1237.
10. Gelman A, King G, Liu C. Not asked and not answered: multiple imputation for multiple surveys (with discussion). *Jounral of the American Statistical Association* 1998; **93**:846–874.
11. Dominici F, Parmigiani G, Reckhow KH, Wolpert RL. Combining information from related regressions. *Journal of Agricultural, Biological, and Environmental Statistics* 1997; **3**:313–332.
12. Rubin DB. *Multiple Imputation for Nonresponse in Sample Surveys.* Wiley, New York, 1987.
13. Mandema JW, Verotta D, Sheiner LB. Building population pharmacokinetic-pharmacodynamic models. *Journal of Pharmacokinetics and Biopharmaceutics* 1992; **20**:511–528.
14. Tanner MA, Wong WH. The calculation of posterior distributions by data augmentation (with discussion). *Journal of the American Statistical Association* 1987; **82**:528–550.
15. Gelfand AE, Smith AFM. Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association* 1990; **85**:398–409.
16. Geman S, Geman D. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 1984; **6**:721–742.
17. Little RTA, Rubin DB. *Statistical Analysis with Missing Data.* Wiley, New York, 1987.
18. Gilks WR, Richardson S, Spiegelhalter DJ. *Markov-Chain Monte-Carlo in Practice.* Chapman & Hall, 1996.
19. Pinheiro JC, Bates DM. Mixed-effects models methods and classes for S and Splus. Downloaded from ftp://ftp.stat.wisc.edu/src/NLME/Unix.
20. Lindstrom MJ, Bates DM. Nonlinear mixed effects models for repeated measures data. *Biometrics* 1990; **46**: 673–687.

