Published in final edited form as:

Biometrics. 2012 September ; 68(3): 943–953. doi:10.1111/j.1541-0420.2011.01719.x.

Bayesian Semiparametric Nonlinear Mixed-Effects Joint Models for Data with Skewness, Missing Responses, and Measurement Errors in Covariates

Yangxin Huang* and Getachew Dagne

Department of Epidemiology & Biostatistics, College of Public Health, University of South Florida, Tampa, Florida 33612, U.S.A

Summary

It is a common practice to analyze complex longitudinal data using semiparametric nonlinear mixed-effects (SNLME) models with a normal distribution. Normality assumption of model errors may unrealistically obscure important features of subject variations. To partially explain between- and within-subject variations, covariates are usually introduced in such models, but some covariates may often be measured with substantial errors. Moreover, the responses may be missing and the missingness may be nonignorable. Inferential procedures can be complicated dramatically when data with skewness, missing values, and measurement error are observed. In the literature, there has been considerable interest in accommodating either skewness, incompleteness or covariate measurement error in such models, but there has been relatively little study concerning all three features simultaneously. In this article, our objective is to address the simultaneous impact of skewness, missingness, and covariate measurement error by jointly modeling the response and covariate processes based on a flexible Bayesian SNLME model. The method is illustrated using a real AIDS data set to compare potential models with various scenarios and different distribution specifications.

Keywords

Bayesian analysis; Covariate measurement errors; Longitudinal data; Missing data; Random-effects models; Skew distributions

1. Introduction

Longitudinal data analysis has attracted considerable research interest and, as a result, a considerable number of statistical modeling and analysis methods have been suggested for analyzing such data with various features. However, there have been few studies on simultaneously accounting for skewness, nonignorable missingness, and covariate measurement errors, which are inherent features of longitudinal data. First, the commonly assumed distribution for model random errors is normal, but this assumption may lack robustness against departures from normality and/or outliers and may obscure important features of between- and within-subject variations of longitudinal data that may exhibit

© 2011, The International Biometric Society

*yhuang@health.usf.edu.

Supplementary Materials

Web Appendices referenced in Sections 1, 3, and 4.1 for multivariate skew distributions, ST–SNLME models in general forms, and WinBUGS code for model I (ST model) are available under the Paper Information link at the *Biometrics* website <http://www.biometrics.tibs.org/>.

departures from symmetry. Thus, analysis under normality may lead to misleading results (Verbeke and Lesaffre, 1996; Sahu, Dey, and Branco, 2003). Second, the validity of inference methods relies on an important requirement that variables are “perfectly” measured. In practice, however, collected data are often far from “perfect.” Although longitudinal studies are designed to collect data from every participant in the study at each assessment time, missing observations in the response are often encountered because of patient dropouts due to drug intolerance or other problems. The missing data are likely to be informative (nonignorable) in the sense that missingness may be related to the missing values. Third, measurement error in covariates is another typical feature of longitudinal data, and ignoring this phenomenon may result in biased statistical inference.

The majority of the statistical literature for the modeling of longitudinal data has focused on the development of models that aim at capturing only specific aspects of the motivating case studies. These include, but are not limited to, a linear mixed-effects (LME) model with a mixture of normal distributions (Verbeke and Lesaffre, 1996); a skew-normal (SN) nonlinear mixed-effects (NLME) model for longitudinal data with skewness and mismeasured covariates (Huang and Dagne, 2011); a semiparametric nonlinear mixed-effects (SNLME) model with a normal distribution in a longitudinal setting with complete data (Wu and Zhang, 2002); and an NLME model with a normal distribution for longitudinal data with censoring and covariate measurement errors (Wu, 2002). However, relatively few studies have been conducted on simultaneously accounting for the biases induced by asymmetry, missing values, and mismeasured covariates. It is not clear how asymmetry, incompleteness, and covariate measurement error of data may interact and simultaneously influence inferential procedures. Statistical inference and analysis complicate dramatically when these three features arise. The goal of this article is to investigate the effects on inference based on a flexible SNLME joint model when these three typical features exist in a longitudinal case study. Specifically, we employed a Bayesian inference approach to jointly investigate the SNLME model with skew-elliptical (SE) distributions, which include SN and skew- t (ST) for a response process, and the LME model with SE distributions for a covariate measurement error process. We offer a flexible conditional modeling strategy characterized by random effects to modulate the missing response data process. It is noted that ST distribution reduces to SN distribution when degrees of freedom are large. Thus, we used ST distribution to develop joint models and associated statistical methodologies, but this easily extends to other SE distributions, including SN distribution. We consider the ST and SN distributions introduced by Sahu et al. (2003), which are suitable for a Bayesian computation and are briefly discussed in Web Appendix A.

The remainder of the article is organized as follows. In Section 2, we describe the data set that motivated this research and investigate specific joint models that include ST–SNLME models for viral load response process, the ST–LME model for CD4 covariate process with measurement error, and the missing response data model. A general form of ST–SNLME models for a response process is presented in Web Appendix B. Section 3 presents the associated Bayesian inference method that simultaneously accounts for skewness, missingness and covariate measurement error. In Section 4, we apply the proposed method to the real data set described in Section 2 and report the analysis results. We conclude the article with discussion in Section 5.

2. Data and Joint Models with the Skew- t Distribution for HIV Dynamics

2.1 Motivating Data Set

The data set that motivated this research is from an AIDS clinical trial study (Acosta et al., 2004). This study consists of 44 HIV-infected patients who were treated with a potent antiretroviral regimen. Although RNA viral load was designed to be measured in copies per

milliliter at study days 0, 7, 14, 28, 56, 84, 112, 140, and 168 of follow-up, it contains missing observations. The number of measurements for each individual varies from four to nine. Six patients dropped out of the study due to drug intolerance or other problems, and 16 patients had missing viral loads at scheduled time points. The exact day of viral load measurement (not predefined study day) was used to compute study day in our analysis. A natural log transformation of viral load was used in the analysis. Covariates such as CD4 cell count and other variables were also measured throughout the study on a similar scheme. The detailed data description can be found in Acosta et al. (2004) and Huang and Dagne (2011).

We can see from the top panel in Figure 1, which shows the pooled viral load and CD4 cell count measurements from all patients, that for this data set to be analyzed in this article, both viral load response (even after ln transformation) and CD4 covariate are highly skewed and, thus, an asymmetric distribution (such as ST or SN) should be more appropriate than a symmetric (normal) distribution. Note that we chose viral load as the dependent variable and considered CD4 cell count as the covariate for the reasons discussed in Web Appendix A by Huang and Dagne (2011) and the fact that our objective in this study was HIV viral dynamics. Figure 1c shows the viral load trajectories after the initiation of an antiretroviral treatment for five randomly selected patients. We see that the viral load trajectories in the initial period followed a clear pattern (a rapid initial decay, called first-phase viral decay) and could be modeled parametrically. After the initial period, however, the viral load trajectories can be quite complicated (a slower decay and some may rebound, called second-phase viral decay). So a parametric model may not be appropriate but, instead, a nonparametric model may be better to capture complicated viral load trajectories.

2.2 Covariate Models with a Skew- t Distribution

This section briefly discusses (CD4) covariate measurement error models with an ST distribution. Various covariate mixed-effects models were investigated in the literature (Wu, 2002; Carrol et al., 2006; and others). The commonly assumed distribution for a measurement error was normal in those statistical covariate models. However, this assumption may lack robustness against departures from normality and/or may violate agreement with the observed data displayed in Figure 1b. We discuss covariate mixed-effects models by considering measurement errors to have an ST distribution. For simplicity, we consider a single time-varying covariate. Denote the number of subjects by n and the number of measurements on the i th subject by n_i . Let z_{ij} be the observed covariate value for individual i at time t_{ij} ($i = 1, \dots, n; j = 1, \dots, n_i$). In the presence of covariate measurement error, we consider the following LME model with the ST distribution to model the CD4 covariate process.

$$z_{ij} = \mathbf{u}_{ij}^T \boldsymbol{\alpha} + \mathbf{v}_{ij}^T \mathbf{a}_i + \varepsilon_{ij} \quad (\equiv z_{ij}^* + \varepsilon_{ij}), \quad (1)$$

$$\varepsilon_i \sim i.i.d. ST_{n_i, \nu_1}(-J(\nu_1)\delta_\varepsilon \mathbf{1}_{n_i}, \sigma_\varepsilon^2 \mathbf{I}_{n_i}, \delta_\varepsilon \mathbf{I}_{n_i}),$$

where $\mathbf{z}_i = (z_{i1}, \dots, z_{in_i})^T$ with z_{ij} being the observed CD4 covariate value for individual i at time t_{ij} , $\mathbf{z}_i^* = (z_{i1}^*, \dots, z_{in_i}^*)^T$ and $\mathbf{z}_{ij}^* = \mathbf{u}_{ij}^T \boldsymbol{\alpha} + \mathbf{v}_{ij}^T \mathbf{a}_i$ may be viewed as the true (but unobservable) covariate value at time t_{ij} , $\mathbf{u}_{ij}$ and $\mathbf{v}_{ij}$ are 1×1 design vectors, $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_I)^T$ and $\mathbf{a}_i = (a_{i1}, \dots, a_{iI})^T$ are unknown population (fixed effects) and individual-specific (random effects) parameter vectors, respectively, $\boldsymbol{\varepsilon}_i = (\varepsilon_{i1}, \dots, \varepsilon_{in_i})^T$ follows a multivariate ST distribution with degrees of freedom ν_1 , unknown scale parameter σ_ε^2 and skewness parameter δ_ε , $J(\nu_1) = (\nu_1/\pi)^{1/2}[\Gamma((\nu_1 - 1)/2)/\Gamma(\nu_1/2)]$, and $\mathbf{1}_{n_i} = (1, \dots, 1)^T$. Note that we specify the skewness matrix being $\delta_\varepsilon \mathbf{I}_{n_i}$; this indicates that we are interested in skewness of overall pooled measurements from all the patients. We assume that $\mathbf{a}_i \sim i.i.d. N_I(0, \boldsymbol{\Sigma}_a)$, where $\boldsymbol{\Sigma}_a$ is

unrestricted covariance matrix. Model (1) may be interpreted as an ST covariate measurement error model that incorporates skewness exhibited in the data.

2.3 Skew- t SNLME Models for HIV Dynamics

Wu and Zhang (2002) and Wu (2002) provided some biological arguments for a two-compartment viral dynamic model where two phases of viral decay are discussed. Thus, we consider an SNLME model for long-term HIV viral dynamics, which may fit the data better than parametric NLME models. Let y_{ij} be the viral load value for individual i at time t_{ij} subject to informative missingness ($i = 1, \dots, n; j = 1, \dots, n_i$) and $\mathbf{y}_i = (y_{i1}, \dots, y_{in_i})^T$. Let $\mathbf{r}_i = (r_{i1}, \dots, r_{in_i})^T$ be a vector of missing response indicator such that $r_{ij} = 1$ if y_{ij} is missing and 0 otherwise. We write $\mathbf{y}_i = (\mathbf{y}_{mis,i}, \mathbf{y}_{obs,i})$, where $\mathbf{y}_{mis,i}$ and $\mathbf{y}_{obs,i}$ are the collections of the missing and observed components of $\mathbf{y}_i$, respectively. For the viral load response process, we consider the following SNLME model, which incorporates possibly mismeasured time-varying CD4 covariate and asymmetric features in viral load response with an ST distribution.

$$\begin{aligned} y_{ij} &= \ln\{e^{p_{i1} - \lambda_{ij1} t_{ij}} + e^{p_{i2} - \lambda_{ij2}(t_{ij})} t_{ij}\} + e_{ij}, \\ p_{i1} &= \beta_1 + b_{i1}, \quad \lambda_{ij1} = \beta_2 + \beta_3 z_{ij}^* + b_{i2}, \\ p_{i2} &= \beta_4 + b_{i3}, \quad \lambda_{ij2}(t_{ij}) = w(t_{ij}) + h_i(t_{ij}), \end{aligned} \quad (2)$$

where y_{ij} is the natural log transformation of the viral load; $\exp(p_{i1}) + \exp(p_{i2})$ is the baseline viral load, and λ_{ij1} and λ_{ij2} are the first- and second-phase viral decay rates, respectively; the vector of model errors

$\mathbf{e}_i = (e_{i1}, \dots, e_{in_i})^T \sim i.i.d. ST_{n_i, \nu_2}(-J(\nu_2)\delta_e \mathbf{1}_{n_i}, \sigma_e^2 \mathbf{I}_{n_i}, \delta_e \mathbf{I}_{n_i})$, which follows a multivariate ST distribution with degrees of freedom ν_2 , unknown scale parameter σ_e^2 , and skewness parameter δ_e ; $w(t)$ and $h_i(t)$ are unknown nonparametric smooth fixed- and random-effects functions, respectively; $h_i(t)$ are independent and identically distributed (i.i.d.) realizations of a zero-mean stochastic process; $\boldsymbol{\beta}_{ij} = (p_{i1}, p_{i2}, \lambda_{ij1}, \lambda_{ij2})^T$ is a vector of individual parameters for the i th subject at time t_{ij} and $\boldsymbol{\beta} = (\beta_1, \beta_2, \beta_3, \beta_4)^T$ is a vector of population parameters; $\mathbf{b}_i^\dagger = (b_{i1}, b_{i2}, b_{i3})^T$ is a vector of random effects.

Model (2) allows the first- and second-phase viral decay rates to change over time. Due to long-term clinical factors, the viral load trajectories can be very complex and may vary over time after the initial viral decay. Thus, for the second-phase viral decay, nonparametric modeling (Wu and Zhang, 2002) may be more appropriate than parametric modeling adopted by Wu (2002) and Huang and Dagne (2011). Therefore, model (2) accommodates not only the skewness nature in viral load data, but also an unknown nonparametric smooth function and expected (but unobservable) CD4 covariate z_{ij}^* rather than observed covariate z_{ij} , which may be mismeasured. The value z_{ij}^* and the unknown nonparametric smooth function are incorporated into the first- and second-phase decay rates, respectively, to capture long-term viral load trajectories with different shapes including viral load rebound. It is noted that the parametric NLME model conducted by Huang and Dagne (2011) is a special case of our model with an SN distribution (2) when model (2) is tailored as follows: (i) the missing response data models are not considered in the joint model in the sense that the missing data in the viral load are missing at random, so that the missing data mechanism is ignored in the analysis; (ii) the unknown nonparametric smooth function for the second-phase decay rate is substituted by a known explicit function, i.e., $w(t_{ij}) = \beta_5 + \beta_6 z_{ij}^*$ and $h(t_{ij}) = b_{i4}$; and (iii) the baseline CD4 values rather than the expected time-varying CD4 values are incorporated into the first-phase viral decay rate λ_{ij1} .

ST-SNLME model (2) is more flexible than commonly used NLME models. To fit model (2), we apply the regression spline method. The working principle is briefly described as follows. The main idea of a regression spline is to approximate $w(t)$ and $h_i(t)$ by using a linear combination of spline basis functions. For instance, $w(t)$ and $h_i(t)$ can be approximated by a linear combination of basis functions $\Psi_p(t) = \{\psi_0(t), \psi_1(t), \dots, \psi_{p-1}(t)\}^T$ and $\Phi_q(t) = \{\phi_0(t), \phi_1(t), \dots, \phi_{q-1}(t)\}^T$, respectively. That is,

$$\begin{aligned} w(t) &\approx w_p(t) = \sum_{l=0}^{p-1} \mu_l \psi_l(t) = \Psi_p(t)^T \mu_p, \\ h_i(t) &\approx h_{iq}(t) = \sum_{l=0}^{q-1} \xi_{il} \phi_l(t) = \Phi_q(t)^T \xi_{iq}, \end{aligned} \quad (3)$$

where μ_p and ξ_{iq} ($q \leq p$ to limit the dimension of random effects for easy implementation) are the unknown vectors of fixed and random coefficients, respectively. Based on the assumption of $h_i(t)$, we can regard ξ_{iq} as i.i.d. realizations of a zero-mean random vector. For our model, we consider natural cubic spline bases with percentile-based knots. To select an optimal degree of regression spline and numbers of knots, i.e., optimal sizes of p and q , the Akaike information criterion (AIC) or the Bayesian information criterion (BIC) is often applied (Wu and Zhang, 2002).

Let $\beta = (\beta^{\dagger T}, \mu_p^T)^T$ and $b_i = (b_i^{\dagger T}, \xi_{iq}^T)^T$ follow $N_{s_1}(\mathbf{0}, \Sigma_b)$, where $s_1 = 3 + q$ and Σ_b is an unstructured covariance matrix. The randomness of the nonparametric mixed effects is transferred to the randomness of the associated coefficients, whereas the nonparametric feature is represented by the basis functions. Thus, for given $\Psi_p(t)$ and $\Phi_q(t)$, substituting $w(t)$ and $h_i(t)$ by their approximations $w_p(t)$ and $h_{iq}(t)$, we can approximate $\lambda_{ij2}(t_{ij})$ using the basis vectors $\Psi_p(t)$ and $\Phi_q(t)$. That is

$$\lambda_{ij2}(t_{ij}) \approx \Psi_p(t)^T \mu_p + \Phi_q(t)^T \xi_{iq}. \quad (4)$$

Thus, SNLME model (2) associated with equation (4) reverts to a parametric NLME model that is similar to those discussed by Huang and Dagne (2011) and Wu (2002).

To allow for informative missing mechanisms in the viral load response, we should assume a missing data model for the missing response mechanism. As discussed in Section 1, we assume that the missing response is related to the random effects a_i and b_i in the response and covariate models. That is, the missingness of the response may depend on the unobservable true underlying response and covariate processes characterized by the random effects. Such a missing data model appears to be of particular interest in HIV viral dynamic studies. For example, we may assume

$$\text{logit}[P(r_{ij}=1|a_i, b_i; \eta)] = \eta_0 + \eta_1 r_{i,j-1} + \eta_2^T a_i + \eta_3^T b_i, \quad (5)$$

where $\eta = (\eta_0, \eta_1, \eta_2^T, \eta_3^T)^T$ is a vector of unknown nuisance parameters. More complicated missing response models can also be considered. Note that the assumed missing response models are not testable based on the observed data. So it is important to carry out sensitivity analyses based on alternative missing data models.

3. Simultaneous Bayesian Inference Method

In a longitudinal study, such as the AIDS study described in Section 2, the longitudinal response and covariate processes are usually connected physically or biologically. Although a simultaneous inference method based on a joint likelihood for the covariate and response data with skewness, incompleteness, and measurement error may be favorable, the computation associated with the joint likelihood inference in such models with the skew distributions for longitudinal data can be extremely intensive and, particularly, may lead to convergence problems (Wu, 2002). Here, we propose a fully Bayesian method for the covariate and response models (1) and (2) to estimate all parameters simultaneously, and we obtain numerical approximations to posterior distributions using Markov chain Monte Carlo (MCMC) methods.

Following discussion in Sahu et al. (2003), to implement an MCMC procedure to models (1) and (2), it can be shown by introducing two $n_i \times 1$ random vectors $\mathbf{w}_{e_i}$ and $\mathbf{w}_{e_i}$ based on the stochastic representation for the ST distribution (see Web Appendix A.1 in detail) that $\mathbf{z}_i$ and $\mathbf{y}_i$ associated with the missing data model (5) can be hierarchically formulated as follows.

$$\begin{aligned} \mathbf{z}_i | \mathbf{a}_i, \mathbf{w}_{e_i}; \alpha, \sigma_1^2, \delta_\varepsilon, \nu_1 &\sim t_{n_i, \nu_1 + n_i}(\mathbf{z}_i^* + \delta_\varepsilon [\mathbf{w}_{e_i} - J(\nu_1 + n_i) \mathbf{1}_{n_i}], \omega_{1i} \sigma_1^2 \mathbf{I}_{n_i}), \\ \mathbf{y}_i | \mathbf{z}_i, \mathbf{r}_i, \mathbf{a}_i, \mathbf{b}_i, \mathbf{w}_{e_i}; \alpha, \beta, \sigma_2^2, \delta_e, \nu_2 &\sim t_{n_i, \nu_2 + n_i}(\mathbf{g}_i(\mathbf{t}_i, \boldsymbol{\beta}_i) + \delta_e [\mathbf{w}_{e_i} - J(\nu_2 + n_i) \mathbf{1}_{n_i}], \omega_{2i} \sigma_2^2 \mathbf{I}_{n_i}), \\ \mathbf{w}_{e_i} &\sim t_{n_i, \nu_1}(\mathbf{0}, \mathbf{I}_{n_i}) I(\mathbf{w}_{e_i} > \mathbf{0}), \\ \mathbf{w}_{e_i} &\sim t_{n_i, \nu_2}(\mathbf{0}, \mathbf{I}_{n_i}) I(\mathbf{w}_{e_i} > \mathbf{0}), \\ \mathbf{r}_{ij} | \mathbf{a}_i, \mathbf{b}_i; \eta &\sim \text{Bernoulli}(P_{ij}), \quad \mathbf{a}_i \sim N_l(\mathbf{0}, \Sigma_a), \\ \mathbf{b}_i &\sim N_{s_1}(\mathbf{0}, \Sigma_b), \end{aligned} \quad (6)$$

where $\omega_{1i} = (\nu_1 + \mathbf{w}_{e_i}^T \mathbf{w}_{e_i}) / (\nu_1 + n_i)$, $\omega_{2i} = (\nu_2 + \mathbf{w}_{e_i}^T \mathbf{w}_{e_i}) / (\nu_2 + n_i)$, $\mathbf{g}_i(\mathbf{t}_i, \boldsymbol{\beta}_i) = (g(t_{i1}, \boldsymbol{\beta}_{i1}), \dots, g(t_{in_i}, \boldsymbol{\beta}_{in_i}))^T$, $g(t_{ij}, \boldsymbol{\beta}_{ij}) = \ln \{ e^{\rho_{1l}} - \lambda_{ij} 2^{t_{ij}} + e^{\rho_{2l}} - \lambda_{ij} 2^{(t_{ij})^{n_j}} \}$, $\mathbf{t}_i = (t_{i1}, \dots, t_{in_i})^T$, $\boldsymbol{\beta}_i = (\boldsymbol{\beta}_{i1}, \dots, \boldsymbol{\beta}_{in_i})^T$, $t_{n_j} \sim \sqrt{(\boldsymbol{\mu}, \mathbf{A})}$ denote the n_j -variate t distribution with parameters $\boldsymbol{\mu}$, $\mathbf{A}$ and degrees of freedom ν , $I(\mathbf{w} > \mathbf{0})$ is an indicator function and $\mathbf{w} = |\boldsymbol{\zeta}|$ with $\boldsymbol{\zeta} \sim t_{n_j} \sim \sqrt{(\mathbf{0}, \mathbf{I}_{n_j})}$, and $P_{ij} = P(r_{ij} = 1 | \mathbf{a}_i, \mathbf{b}_i; \eta)$. Note that an important advantage of the above representations based on the hierarchical models under a Bayesian framework is that they allow one to easily implement the methods using the freely available WinBUGS software (Lunn et al., 2000) and that the computational effort for the model with an ST distribution is almost equivalent to that of the model with a t -distribution.

Let $\theta = \{\alpha, \beta, \eta, \sigma_1^2, \sigma_2^2, \Sigma_a, \Sigma_b, \nu_1, \nu_2, \delta_\varepsilon, \delta_e\}$ be the collection of unknown parameters in models (1), (2), and (5). Under the Bayesian framework, we next need to specify prior distributions for all the unknown parameters in these three models, as follows:

$$\begin{aligned} \alpha &\sim N_l(\boldsymbol{\tau}_1, \mathbf{A}_1), \quad \sigma_1^2 \sim IG(\omega_1, \omega_2), \\ \Sigma_a &\sim IW(\boldsymbol{\Omega}_1, \rho_1), \quad \delta_\varepsilon \sim N(0, \gamma_1), \\ \beta &\sim N_{s_2}(\boldsymbol{\tau}_2, \mathbf{A}_2), \quad \sigma_2^2 \sim IG(\omega_3, \omega_4), \\ \Sigma_b &\sim IW(\boldsymbol{\Omega}_2, \rho_2), \quad \delta_e \sim N(0, \gamma_2), \\ \nu_1 &\sim \text{Exp}(\nu_{10}) I(\nu_1 > 3), \quad \nu_2 \sim \text{Exp}(\nu_{20}) I(\nu_2 > 3), \\ \eta &\sim N_{s_3}(\boldsymbol{\tau}_3, \mathbf{A}_3), \end{aligned} \quad (7)$$

where $s_2 = 4 + p$, $s_3 = 2 + s_1 + l$, the mutually independent inverse gamma (IG), normal (N), exponential (Exp) and inverse Wishart (IW) prior distributions are chosen to facilitate computations. The super-parameter matrices $\mathbf{A}_1$, $\mathbf{A}_2$, $\mathbf{A}_3$, $\boldsymbol{\Omega}_1$, and $\boldsymbol{\Omega}_2$ can be assumed to be diagonal for convenient implementation.

Let the observed data $\mathcal{D} = \{(\mathbf{y}_{obs,i}, \mathbf{z}_i, \mathbf{r}_i), i = 1, \dots, n\}$. Let $f(\cdot)$, $f(\cdot|\cdot)$ and $\pi(\cdot)$ be a generic density function, a conditional density function, and a prior density function, respectively. We assume that $\mathbf{a}, \mathbf{\beta}, \boldsymbol{\eta}, \sigma_1^2, \sigma_2^2, \boldsymbol{\Sigma}_a, \boldsymbol{\Sigma}_b, \nu_1, \nu_2, \delta_e$, and δ_e are independent of each other, i.e., $\pi(\boldsymbol{\theta}) = \pi(\boldsymbol{\alpha})\pi(\boldsymbol{\beta})\pi(\boldsymbol{\eta})\pi(\sigma_1^2)\pi(\sigma_2^2)\pi(\sum_a)\pi(\sum_b)\pi(\nu_1)\pi(\nu_2)\pi(\delta_e)\pi(\delta_e)$. After we specify the models for the observed data and the prior distributions for the unknown model parameters, we can make statistical inference for the parameters based on their posterior distributions under the Bayesian framework. When the missing responses are nonignorable, we need to assume only a possible missing data model and then incorporate it into likelihood. Thus, the joint posterior density of $\boldsymbol{\theta}$ based on the observed data $\mathcal{D}$ can be given by

$$\begin{aligned} f(\boldsymbol{\theta}|\mathcal{D}) &\propto \left\{ \prod_i^n \int \int \int f(\mathbf{y}_i|\mathbf{z}_i, \mathbf{r}_i, \mathbf{a}_i, \mathbf{b}_i, \mathbf{w}_{e_i}; \boldsymbol{\alpha}, \boldsymbol{\beta}, \sigma_2^2, \delta_e, \nu_2) \right. \\ &\quad \times f(\mathbf{z}_i|\mathbf{a}_i, \mathbf{w}_{e_i}; \boldsymbol{\alpha}, \sigma_2^2, \delta_e, \nu_1) \\ &\quad \times f(\mathbf{w}_{e_i}|\mathbf{w}_{e_i} > \mathbf{0})f(\mathbf{w}_{e_i}|\mathbf{w}_{e_i} > \mathbf{0})f(\mathbf{r}_i|\mathbf{a}_i, \mathbf{b}_i; \boldsymbol{\eta}) \\ &\quad \times f(\mathbf{a}_i)f(\mathbf{b}_i)d\mathbf{y}_{mis,i}d\mathbf{a}_id\mathbf{b}_i \Big\} \pi(\boldsymbol{\theta}) \\ &= \left\{ \prod_i^n \int \int \int f(\mathbf{y}_{obs,i}|\mathbf{z}_i, \mathbf{r}_i, \mathbf{a}_i, \mathbf{b}_i, \mathbf{w}_{e_i}; \boldsymbol{\alpha}, \boldsymbol{\beta}, \sigma_2^2, \delta_e, \nu_2) \right. \\ &\quad \times f(\mathbf{z}_i|\mathbf{a}_i, \mathbf{w}_{e_i}; \boldsymbol{\alpha}, \sigma_2^2, \delta_e, \nu_1) \\ &\quad \times f(\mathbf{w}_{e_i}|\mathbf{w}_{e_i} > \mathbf{0})f(\mathbf{w}_{e_i}|\mathbf{w}_{e_i} > \mathbf{0})f(\mathbf{r}_i|\mathbf{a}_i, \mathbf{b}_i; \boldsymbol{\eta}) \\ &\quad \times f(\mathbf{a}_i)f(\mathbf{b}_i)d\mathbf{a}_id\mathbf{b}_i \Big\} \pi(\boldsymbol{\theta}). \end{aligned} \quad (8)$$

Note that the missing response $\mathbf{y}_{mis,i}$ is integrated out, which is not possible for missing data models where the missingness is directly related to the missing response values (Ibrahim, Chen, and Lipsitz, 2001). In general, the integrals in (8) are of high dimension and do not have closed form. Analytic approximations to the integrals may not be sufficiently accurate. Therefore, it is prohibitive to directly calculate the posterior distribution of $\boldsymbol{\theta}$ based on the observed data. As an alternative, the MCMC procedure can be used to sample from the posterior distributions, based on (8), using the Gibbs sampler along with the Metropolis–Hastings (M–H) algorithm.

4. Data Analysis

4.1 Model Implementation

The ST–SNLME model (2) in Section 2, which specifies the second-phase viral decay rate λ_{ij2} nonparametrically, appears to provide a reasonable fit to the observed data. We consider linear combinations of natural cubic splines with percentile-based knots to approximate the nonparametric functions $w(t)$ and $h_i(t)$. Following the study in Wu and Zhang (2002) for likelihood inference under the model with a normal distribution, we set $\psi_0(t) = \varphi_0(t) \equiv 1$ and take the same natural cubic splines in the approximations (3) with q p (to limit the dimension of random effects). The values of p and q are determined by the AIC/BIC criteria. The AIC/BIC values are evaluated based on the standard normal model with various (p, q) combinations $(p, q) = \{(1, 1), (2, 1), (2, 2), (3, 1), (3, 2), (3, 3)\}$, which suggest the following function for $\lambda_{ij2}(t)$ with $p = 3$ and $q = 1$ in model (2):

$$\lambda_{ij2}(t_{ij}) \approx \beta_5 + \beta_6\psi_1(t_{ij}) + \beta_7\psi_2(t_{ij}) + b_{i4}. \quad (9)$$

Thus, we have population parameter vector $\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_7)^T$ and individual random effects $\mathbf{b}_i = (b_{i1}, b_{i2}, b_{i3}, b_{i4})^T$ following $N_4(\mathbf{0}, \boldsymbol{\Sigma}_b)$.

Figure 1d shows the CD4 trajectories of five randomly selected patients. It is known that CD4 is often measured with substantial error, so it is reasonable to assume that λ_{ij1} in model (2) is related to the true but unobserved CD4 values rather than the observed but mismeasured CD4 values. To obtain reliable estimates of the viral dynamic parameters, which can be used to evaluate the anti-HIV treatment, it is important to simultaneously address measurement errors in the CD4 covariate and missing data in the viral load response. In the absence of a theoretical rationale for the CD4 trajectories, we consider empirical polynomial LME models for the CD4 process and choose the best model based on AIC and BIC values. Specifically, we consider the covariate model (1) with

$u_{ij} = v_{ij} = (1, t_{ij}, \dots, t_{ij}^{l-1})^T$ and focus on linear ($l = 2$), quadratic ($l = 3$) and cubic ($l = 4$) polynomials. The resulting AIC (BIC) values are 799.03 (821.74), 703.56 (744.42), and 766.18 (782.08), respectively. Thus, we adopted the following quadratic polynomial LME model for the CD4 process:

$$z_{ij} = (\alpha_1 + a_{i1}) + (\alpha_2 + a_{i2})t_{ij} + (\alpha_3 + a_{i3})t_{ij}^2 + \varepsilon_{ij}, \quad (10)$$

where $z_{ij}^* = (\alpha_1 + a_{i1}) + (\alpha_2 + a_{i2})t_{ij} + (\alpha_3 + a_{i3})t_{ij}^2$, $\alpha = (\alpha_1, \alpha_2, \alpha_3)^T$ is a vector of population (fixed effects) parameters, $\varepsilon_i = (\varepsilon_{i1}, \dots, \varepsilon_{in_i})^T \sim ST_{n_i, \nu_1}(-J(\nu_1)\delta_\varepsilon \mathbf{1}_{n_i}, \sigma_\varepsilon^2 \mathbf{I}_{n_i}, \delta_\varepsilon \mathbf{I}_{n_i})$ and individual-specific random effects $\mathbf{a}_i = (a_{i1}, a_{i2}, a_{i3})^T \sim N_3(\mathbf{0}, \Sigma_a)$. In addition, to avoid too small or large estimates that may be unstable, we standardize the time-varying covariate CD4 cell counts (each CD4 value is subtracted by 375.46 and divided by 228.57) and rescale the original time t (in days) so that the time scale is between 0 and 1.

In AIDS studies, some patients may drop out early due to drug side effects or other problems, although some of these dropout patients may return to the study at a later time. In such studies, patient dropouts are often informative (i.e., they may have slower viral decay in comparison with the remaining patients, see Figure 1c). These patient dropouts are likely to depend on the (unobservable) true viral load and CD4 cell count trajectories, which are characterized by the random effects in the models, because the observed values are measured with errors. For example, patients with slower viral decay or slower CD4 increase or higher (lower) initial true viral load (CD4) levels may be more likely to drop out or miss scheduled visits than other patients. So, the missing data may be nonignorable, and a possible missing data model must be incorporated in the analysis to avoid biased results. We consider the following possible nonignorable missing data model for response (denoted by *NIM*):

$$\text{logit}[P(r_{ij}=1|\mathbf{a}_i, \mathbf{b}_i; \eta)] = \eta_0 + \eta_1 r_{i,j-1} + \eta_2 a_{i1} + \eta_3 a_{i2} + \eta_4 a_{i3} + \eta_5 b_{i2} + \eta_6 b_{i4}, \quad (11)$$

which links the missingness to the random effects that characterize the true response and covariate processes and may be reasonable in practice. Note that the nonignorable model (11) is not testable based on the observed data, so we should consider alternative missing data models to perform sensitivity analysis. In practice, it is also likely that the missingness may be related to the values being missing. Thus, we have considered the alternative nonignorable model $\text{logit}[P(r_{ij}=1; \boldsymbol{\eta})] = \boldsymbol{\eta}_0 + \boldsymbol{\eta}_1 r_{i,j-1} + \boldsymbol{\eta}_2 z_{ij} + \boldsymbol{\eta}_3 y_{ij}$. The results are similar to those based on (11), so we omit those results here to save space. For comparison, we also consider the following ignorable missing data model for response (denoted by *IM*):

$$\text{logit}[P(r_{ij}=1; \eta)] = \eta_0 + \eta_1 y_{id}, \quad d \leq j, \quad (12)$$

where y_{id} ($d \leq j$) is the last-observed response for individual i .

As shown in Figure 1, the histograms of viral load in ln scale and CD4 cell count clearly indicate their asymmetric feature. Thus, it seems plausible to fit models with a skew distribution to the data sets. In addition, we note that Huang and Dagne (2011) investigated the NLME model assuming both model errors and random effects to have the SN distribution and the modeling results based on the SN distribution for random effects indicated no significant difference from those based on the normal distribution for random effects. With this information, the following three statistical models with different distribution specifications of model errors for both the response model (2) in conjunction with (9) and the missing data model (11) for viral dynamics and the covariate model (10) for CD4 process were employed to compare their performance:

- **Model I:** A model with the independent multivariate ST distribution of random errors for both the response model (2) and the covariate model (10).
- **Model II:** A model with the independent multivariate SN distribution of random errors for both the response model (2) and the covariate model (10).
- **Model III:** A model with the independent multivariate normal distribution of random errors for both the response model (2) and the covariate model (10).

We investigated the following four scenarios. First, because a normal distribution is a special case of an SN distribution when the skewness parameter is zero, while the ST distribution reduces to the SN distribution when the degree of freedom is large, we investigated how an asymmetric (SN or ST) distribution for model error contributes to modeling results and parameter estimation in comparison with a symmetric (normal) distribution. Second, we estimated the model parameters by using the “naive” method (denoted by NM), which ignores measurement error in CD4 covariate and missing data in viral load response. That is, the naive method uses only the observed CD4 values z_{ij} rather than unobservable CD4 values z_{ij}^* in the response model (2) without the missing viral load data model incorporated into model fitting. We used it as a comparison to the joint modeling (JM) approach proposed in Section 2. This comparison attempted to investigate how the measurement errors in CD4 and missing data in viral load contribute to modeling results. Third, the nonignorable missing data model (11) was replaced by the ignorable missing data model (12) for inference to investigate how missing response data influence modeling results. Finally, we also considered the *parametric* NLME model (denoted by PJM), obtained by setting $w(t_{ij}) = \beta_5 + \beta_6 z_{ij}^*$ and $h(t_{ij}) = b_{i4}$ in the model (2) based on the JM approach, to evaluate how the SNLME model (2) contributes to the modeling results in comparison with the *parametric* NLME model, which includes the model conducted by Huang and Dagne (2011) as a special case.

To carry out the Bayesian inference, we specified the values of the hyperparameters in the prior distributions. In the Bayesian analysis, we specified only the priors at the population level. We took weakly informative prior distribution for the parameters in the joint models. In particular, (i) fixed effects were taken to be independent normal distributions $N(0, 100)$ for each component of the population parameter vectors α , β , and η ; (ii) for the scale parameters σ_1^2 and σ_2^2 we assume a limiting noninformative inverse gamma prior distribution, $IG(0.01, 0.01)$ so that the distribution has mean 1 and variance 100; (iii) the priors for the variance–covariance matrices of the random effects Σ_a and Σ_b were taken to be inverse Wishart distributions $IW(\Omega_1, \rho_1)$ and $IW(\Omega_2, \rho_2)$ with covariance matrices $\Omega_1 = \text{diag}(0.01, 0.01, 0.01)$, $\Omega_2 = \text{diag}(0.01, 0.01, 0.01, 0.01)$ and $\rho_1 = \rho_2 = 4$, respectively; (iv) the degrees of freedom parameters ν_1 and ν_2 followed truncated exponential distribution with $\nu_{10} = \nu_{20} = 0.5$; and (v) for each of the skewness parameters δ_e and δ_s , we chose independent normal distribution $N(0, 100)$.

The MCMC sampler was implemented using WinBUGS (Lunn et al., 2000). The code for model I is available in Web Appendix C. The MCMC scheme for drawing samples from the posterior distributions of all parameters in the joint models was obtained by iterating between the following two steps: (i) Gibbs sampler is used to update $\alpha, \beta, \eta, \sigma_1^2, \sigma_2^2, \Sigma_a, \Sigma_b, \nu_1, \nu_2, \delta_e$, and δ_e ; (ii) we update b_i and a_i ($i = 1, 2, \dots, n$) using the M-H algorithm. After collecting the final MCMC samples, we drew statistical inference for the unknown parameters. Specifically, we were interested in the posterior means and quantiles. See Huang, Liu, and Wu (2006) and Lunn et al. (2000) for detailed discussions of the Bayesian modeling approach and the implementation of the MCMC procedures, including the choice of the hyperparameters, the iterative MCMC algorithm, the choice of proposal density related to M-H sampling, sensitivity analysis, and convergence diagnostics. When the MCMC procedure was applied to the actual clinical data, convergence of the generated samples was assessed using standard tools within WinBUGS (such as trace plots). After convergence was achieved, one long chain was run that may be more efficient with the following considerations. We proposed that, after an initial number of 50,000 burn-in iterations, every 20th MCMC sample be retained from the next 200,000. Thus, we obtained 10,000 samples of targeted posterior distributions of the unknown parameters for statistical inference. The computational burden for fitting such highly nonlinear joint models via MCMC procedure was reasonable. For example, to fit model II, it took about 4.0 hours on a Window workstation with Intel Core 2 Quad CPU @ 2.83GHz and RAM of 8.0 GB.

4.2 Comparison of Modeling Results

The Bayesian JM approach based on the specific SNLME model with different model distribution specifications was used to fit the data. From the model fitting results, we have seen that, in general, all the models provided a reasonably good fit to the observed data for most patients in our study, although the fitting for a few patients (<8%) was not completely satisfactory due to unusual data fluctuation patterns for these patients, particularly for model III. To assess the goodness of fit of the proposed models, the diagnosis plots of the observed values versus the fitted values (top panel) and ST, SN, or normal Q-Q plots (bottom panel) from models I, II, and III are presented in Figure 2. It can be seen from Figure 2 (top panel) that the model where the random error is assumed to have the ST and SN distributions gave a similar performance, but they provided better fit to observed data compared with the model where the random error is assumed to have the normal distribution. This result can also be explained by examining the Q-Q plots of the residuals (bottom panel) that all show the existence of outliers, but it is clearly seen that models I and II only have few outliers and thus give better goodness of fit to the data than that of model III. Further, model II does a better job in accounting for skewness. This finding is further confirmed by their standardized residual sum of squares, which are 68.51 (ST model), 17.58 (SN model), and 261.20 (normal model).

The population posterior mean, the corresponding standard deviation (SD), and 95% credible interval for fixed-effects parameters based on the three models with two methods (JM approach and naive method) are presented in Tables 1 and 2. The following findings are observed for estimated results of parameters. In the response model (2), the findings, particularly for the fixed effects (β_2, β_3), which are parameters of the first-phase viral decay rate λ_1 , show that these estimates are statistically significant for all of three models because the 95% credible intervals do not contain zero. Nevertheless, for the estimate of the coefficient of CD4 covariate β_3 , its estimate is significantly positive. This means that CD4 has a significantly positive effect on the first-phase viral decay rate, suggesting that the CD4 covariate may be an important predictor of the first-phase viral decay rate during the treatment. Further, the estimate of β_3 based on model II is larger than those based on models I and III, and there are marked differences in posterior means of the scale parameter σ_2^2 (0.06

for models I and II, and 1.28 for model III) across the three models considered. The estimated values of the scale parameter for both models I and II are much smaller than that of model III because the former models take into account skewness of the data while the latter does not. The estimates of the skewness parameter (δ_e) of models I and II are, respectively, 1.12 with 95% credible intervals (0.84, 1.33) and 1.75 with 95% credible intervals (1.47, 1.95). This finding suggests that there is a significantly positive skewness in the data and confirms the fact that the distribution of the original data is skewed even after taking $\ln$ transformation (see Figure 1a). Thus, incorporating a skewness parameter in the modeling of the data is recommended. Furthermore, the estimate (0.84) of the skewness parameter based on model I with the ST distribution is less than that (1.75) based on model II with the SN distribution. This may be because an additional parameter ν_2 for heaviness in the tails was estimated to be 3.79 trading-off the effect of skewness.

For parameter estimates of the CD4 covariate model (10), the estimates of the coefficient α_2 based on the three models are slightly different but significantly positive, whereas the estimates of α_1 and α_3 are not statistically significant. This finding suggests that there is a positive linear relation between CD4 cell count and measurement time. There are quite large differences in the posterior means of the scale parameter σ_1^2 : 0.02 for model I, 0.09 for model II, and 0.13 for model III. The estimated values of the scale parameter for models I and II are smaller than that of model III because the former models take into account skewness of the data while the latter does not. Furthermore, model I has the least scale parameter because it considers heaviness in tails in addition to skewness.

For selecting the best model that fits the data adequately, a Bayesian selection criterion, known as deviance information criterion (DIC) suggested by Spiegelhalter et al. (2002), is used. As with other model selection criteria, we caution that DIC is not intended for identification of the “correct” model, but rather merely as a method of comparing a collection of alternative formulations. As one referee suggested, as an alternative, we also evaluate expected predictive deviance (EPD) for model formulated by $EPD = E\{\sum_{i,j} (y_{rep,ij} - y_{obs,ij})^2\}$ comparison, where the predictive value $y_{rep,ij}$ is a replicate of the observed $y_{obs,ij}$ and the expectation is taken over the posterior distribution of the model parameters θ (see Gelman et al., 2003, in detail). This criterion chooses the model where the discrepancy between predictive values and observed values is the lowest. We calculate estimated DICs using the JM approach based on models I, II, and III, which are 1368.9, 572.8, and 2599.1, respectively. As mentioned previously, it is difficult to tell which model is correct, only which one fits the data better. Furthermore, the model that fits the data better may be more accurate to describe the mechanisms of HIV infection and the CD4 changing process, so it better informs patient treatment. Therefore, based on the DIC, the results indicate that model II is the best-fitting model and model I is next, supporting the contention of a departure from normality. This finding is confirmed by the results of the EPD values (see Table 2). These results are consistent with those in diagnosis of the goodness of fit displayed in Figure 2, indicating that model II performs best. In summary, our results may suggest that it is important to assume a skew distribution for the viral load response and CD4 covariate models to achieve reliable results, in particular if the data exhibit skewness. Based on these observations, we will further report our results in detail only for the best model II below.

4.3 Estimation Results Based on Model II

To investigate how the measurement errors in CD4 and missing data in viral load contribute to modeling results, we employed the naive method based on model II, which ignores measurement errors in CD4 and missing data in viral load, to estimate the model parameters. It can be seen from Tables 1 and 2 that there are important differences in the estimates for the parameters β_2 and β_3 , which are directly associated with whether or not ignoring

potential CD4 measurement errors for inference. The naive method may produce biased estimates and may substantially underestimate the covariate CD4 effect (β_3). The estimated SD for the CD4 effect (β_3) using the JM method is at least twice larger than that using the naive method, probably because the JM method incorporates the variation from fitting the CD4 process. The difference of the naive estimates and the JM estimates, due to whether or not potential CD4 measurement errors and missing data in viral load are ignored for inference indicates that CD4 measurement errors and missing data in viral load cannot be ignored in the analysis. We also obtained estimated DIC values with 572.8 (JM) and 635.7 (NM) based on model II, respectively. We can see from the estimated DIC values that the JM approach performs better in comparison with the naive method. Thus, it is important to take the CD4 measurement errors and the missing data in viral load into account when collected data are “imperfectly” measured.

The estimates from the nonignorable and ignorable missing data models in Table 1 appear to be different, especially for our parameters of interest associated with the decay rates λ_{ij1} and λ_{ij2} . This suggests that the ignorable missing data model may lead to some biased estimation and, thus, reliable estimation must incorporate a reasonable non-ignorable missing response data model. Although some estimates in Table 1 are not statistically significant, the values of the estimates may still provide useful information about viral load and CD4 trajectories. Further, the estimates (SD in parentheses) of parameters η in the nonignorable missing data model (11) based on model II are $\hat{\eta} = (0.16(0.09), 1.09(0.05), -0.07(0.10), -0.22(0.07), 0.02(0.06), 0.14(0.09), \text{ and } 0.06(0.08))^T$. This shows that the missingness of response is strongly associated with the previous missing status and the random effects that characterize the viral load and CD4 trajectories. We conducted overall hypothesis testing with $H_0: \eta_1 = \dots = \eta_6 = 0$ using the Wald-type test statistic, which approximately follows χ^2 distribution with six degrees of freedom under H_0 . The test statistic leads to a p -value of 0.03, indicating that the missing viral load data are nonignorable (or informative).

We also considered the JM approach for the *parametric* NLME model, denoted by PJM in comparison with the semi-parametric NLME model developed in this article. We found that the parametric NLME model may lead to possibly biased estimates (see, e.g., estimates of β) in comparison with the semiparametric NLME model. Thus, the semiparametric NLME model offers possibly more reliable estimates of some parameters, which may lead to better understanding of the HIV viral dynamics and the treatment efficacy.

The estimated results based on JM approach (model II) indicate that the population CD4 trajectory may be approximated by the quadratic polynomial $\hat{z}(t) = 228.57(-0.04 + 0.63t - 0.29t^2) + 375.46$, where $\hat{z}(t)$ is in the original CD4 scale. The estimated population first- and second-phase viral decay rates are $\hat{\lambda}_1 = 28.4 + 13.4 \hat{z}(t)$ and $\hat{\lambda}_2(t) = -2.83 - 0.26 \psi_1(t) - 0.55 \psi_2(t)$. Thus, the population viral load process may be approximated by $\hat{V}(t) = \exp(5.25 - \hat{\lambda}_1 t) + \exp(0.41 - \hat{\lambda}_2(t)t)$. Because the first-phase viral decay rate (λ_1) is significantly associated with the true CD4 values (due to the statistically significant estimate of β_3), this suggests that the viral load change $V(t)$ may be significantly associated with the true CD4 values. Note that, although the true association described above may be complicated, the simple approximation considered here may provide a rough guidance and point to further research.

5. Discussion

Modeling skewness by modifying well-known distributions is a topic that has received much attention over the past years (Sahu et al., 2003; Arellano-Valle, Bolfarine, and Lachos, 2007; Ho and Lin, 2010; Huang and Dagne, 2011). In the presence of the three typical features of

data—skewness, missing value in response, and measurement error in covariate, we propose a robust approach to a Bayesian SNLME joint model based on the ST and SN distributions as a powerful tool to analyze longitudinal data with asymmetric, incomplete, and mismeasured natures. The proposed methods enhance modeling flexibility, thereby allowing practitioners to analyze longitudinal data under a wide variety of considerations. In addition, the proposed JM approach can be easily implemented using the publicly available WinBUGS package. This makes our approach quite powerful and accessible to practicing statisticians in the fields.

This article considers three statistical models—ST, SN, and normal distributions-based models. We compared their performance by examining the residual Q–Q plots and residual sum of squares for the assessment of goodness of fit and by using DIC and EPD for the model selection. It is shown from our results that all four criteria achieve a consistent conclusion. It is noted that Celeux et al. (2006) investigated and compared different DIC constructions for missing data models. A further study to evaluate these various extensions to DIC used here may be warranted, but it will require additional effort due to the nature of the complicated nonlinear models with skew distributions whose associated likelihood function cannot be easily derived; thus, it is beyond the scope of this article. We hope to report these results in the near future.

The basic tool for investigating model uncertainty is sensitivity analysis. That is, we simply make reasonable modifications to the assumptions in question, recompute the posterior quantities of interest, and see whether they have changed in a way that significantly affects the interpretations or conclusions. If the results are robust against the suspected assumptions, we can report the results with confidence. However, if the results are sensitive to the assumptions, we can choose to communicate the sensitivity results and interpret the results with caution (Gelman et al., 2003). As the associate editor and one referee suggested, we performed a sensitivity analysis to examine the dependence of parameter estimates on the prior distributions. As an example, we implemented the MCMC sampling scheme and monitored several independent MCMC runs, starting from different super-parameter values in $IG(\cdot, \cdot)$ prior distribution. The results follow the same patterns as those presented in this article. The conclusions of our analysis remain unchanged (data not shown due to space limitations).

In summary, for this particular data set based on model II, our data analysis indicates that for reliable estimation of HIV dynamic parameters, we should simultaneously address measurement error, asymmetry, and missing data. We found that it is important to take the CD4 measurement errors and viral load missing data into account, indicating that the JM approach outperforms the naive method; the missing viral load data are likely to be nonignorable and, thus, estimates under a nonignorable missing data model may be more reliable than those under an ignorable missing data model; the semiparametric NLME model may fit the observed data better than the parametric NLME model.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors gratefully acknowledge the editor, an associate editor, and two anonymous referees for their insightful comments and constructive suggestions that led to a marked improvement of the article. This research was partially supported by NIAID grant AI080338 to YH, and NIMH grant R01MH040859-22 to GD.

References

- Acosta EP, Wu H, Walawander A, Eron J, Pettinelli C, Yu S, Neath D, Ferguson E, Saah AJ, Kuritzkes DR, Gerber JG. Comparison of two indinavir/ritonavir regimens in treatment-experienced HIV-infected individuals. *Journal of AIDS*. 2004; 37:1358–1366. [PubMed: 15483465]
- Arellano-Valle RB, Bolfarine H, Lachos VH. Bayesian inference for skew-normal linear mixed models. *Journal of Applied Statistics*. 2007; 34:663–682.
- Carroll, RJ.; Ruppert, D.; Stefanski, LA.; Crainiceanu, CM. *Measurement Error in Nonlinear Models: A Modern Perspective*. 2. London: Chapman and Hall; 2006.
- Celeux G, Forbes F, Robert C, Titterton M. Deviance information criteria for missing data models. *Bayesian Analysis*. 2005; 1:651–674.
- Gelman, A.; Carlin, JB.; Stern, HS.; Rubin, DB. *Bayesian Data Analysis*. London: Chapman and Hall; 2003.
- Ho H, Lin T. Robust linear mixed models using the skew t distribution with application to schizophrenia data. *Biometrical Journal*. 2010; 52:449–469. [PubMed: 20680971]
- Huang Y, Dagne G. A Bayesian approach to joint mixed-effects models with a skew-normal distribution and measurement errors in covariates. *Biometrics*. 2011; 67:260–269. [PubMed: 20486924]
- Huang Y, Liu D, Wu H. Hierarchical Bayesian methods for estimation of parameters in a longitudinal HIV dynamic system. *Biometrics*. 2006; 62:413–423. [PubMed: 16918905]
- Ibrahim JG, Chen MH, Lipsitz SR. Missing responses in generalized linear mixed models when the missing data mechanism is nonignorable. *Biometrika*. 2001; 88:551–564.
- Lunn DJ, Thomas A, Best N, Spiegelhalter D. WinBUGS—a Bayesian modelling framework: Concepts, structure, and extensibility. *Statistics and Computing*. 2000; 10:325–337.
- Sahu SK, Dey DK, Branco MD. A new class of multivariate skew distributions with applications to Bayesian regression models. *The Canadian Journal of Statistics*. 2003; 31:129–150.
- Spiegelhalter DJ, Best NG, Carlin BP, Van der Linde A. Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society, Series B*. 2002; 64:583–639.
- Verbeke G, Lesaffre E. A linear mixed-effects model with heterogeneity in random-effects population. *Journal of the American Statistical Association*. 1996; 91:217–221.
- Wu L. A joint model for nonlinear mixed-effects models with censoring and covariates measured with error. *Journal of the American Statistical Association*. 2002; 97:955–964.
- Wu H, Zhang JT. The study of long-term HIV dynamics using semi-parametric non-linear mixed-effects models. *Statistics in Medicine*. 2002; 21:3655–3675. [PubMed: 12436462]

Figure 1.

Histograms (top panel) of viral load measured from RNA levels (in natural ln scale) and standardized CD4 cell count in plasma for 44 patients in an AIDS clinical trial study. Profiles (bottom panel) of viral load (response) in ln scale and CD4 cell count (covariate) for five randomly selected patients.

Figure 2. Goodness of fit: Observed values versus fitted values of $\ln(\text{RNA})$ (top panel), and ST, SN, and normal Q-Q plots with line (bottom panel).

Table 1

Summary of estimated posterior mean (PM) of population (fixed effects) parameters, the corresponding SD, and lower limit (L_{CI}) and upper limit (U_{CI}) of 95% equal-tail credible interval (CI) based on naive method (NM) and JM approach for semiparametric NLME model and PJM in conjunction with nonignorable missing (NIM) or ignorable missing (IM) response data

Missing data	Method	Model		β_1	β_2	β_3	β_4	β_5	β_6	β_7	α_1	α_2	α_3
NIM	JM	I	PM	5.86	30.1	9.43	1.89	-0.38	-1.31	0.96	-0.78	0.64	-0.30
			L_{CI}	5.23	24.4	3.64	1.29	-2.37	-3.31	-18.5	-1.11	0.26	-0.79
			U_{CI}	6.47	37.8	16.6	3.29	2.92	0.41	20.9	-0.47	1.05	0.14
	JM	II	SD	0.32	3.33	3.29	0.46	1.16	0.93	10.0	0.15	0.20	0.23
			PM	5.25	28.4	13.4	0.41	-2.83	-0.26	-0.55	-0.04	0.63	-0.29
			L_{CI}	4.60	17.3	2.55	-2.40	-10.5	-3.32	-20.4	-1.02	0.15	-0.76
NIM	JM	III	U_{CI}	5.95	43.7	26.9	2.28	1.71	5.40	18.8	0.52	1.11	0.24
			SD	0.35	6.92	7.62	1.50	3.84	2.05	9.98	0.50	0.24	0.26
			PM	8.11	30.3	9.19	3.85	-1.95	-0.30	-0.41	-0.20	0.61	-0.23
	NM	II	L_{CI}	7.59	23.1	1.54	2.59	-6.03	-3.49	-19.9	-0.47	0.05	-0.79
			U_{CI}	8.62	39.2	17.4	4.99	2.07	2.90	19.1	0.07	1.16	0.35
			SD	0.26	4.14	4.04	0.62	2.05	1.65	9.93	0.14	0.28	0.29
NIM	PJM	II	PM	5.34	26.9	9.98	0.65	-2.73	0.40	0.77	-	-	-
			L_{CI}	4.76	21.0	4.76	-0.48	-5.28	-1.57	-19.2	-	-	-
			U_{CI}	5.98	32.8	16.2	1.64	-0.30	2.24	20.3	-	-	-
	PJM	II	SD	0.32	3.09	2.91	0.58	1.45	1.01	10.0	-	-	-
			PM	5.19	25.0	8.00	0.46	-3.00	0.99	-	-0.82	0.58	-0.21
			L_{CI}	4.61	19.4	2.56	-0.35	-4.37	0.27	-	-1.13	0.08	-0.77
IM	JM	II	U_{CI}	5.82	30.4	14.5	1.46	-1.45	1.83	-	-0.48	1.09	0.34
			SD	0.32	2.68	2.99	0.45	0.72	0.40	-	0.17	0.27	0.29
			PM	5.10	24.4	8.16	-0.05	-4.51	1.30	0.71	-0.47	0.64	-0.26
	NM	II	L_{CI}	4.48	18.4	0.77	-1.28	-7.56	-1.83	-18.7	-1.16	0.16	-0.80
			U_{CI}	5.66	33.8	19.6	1.26	-0.62	3.89	20.4	0.51	1.15	0.27
			SD	0.32	3.87	4.55	0.69	2.16	1.69	9.98	0.49	0.25	0.27

Table 2

Summary of estimated posterior mean (PM) for parameters of scale, skewness, and degree of freedom, and corresponding SD and lower limit (L_{CI}) and upper limit (U_{CI}) of 95% equal-tail credible interval (CI) as well as DIC and EPD values based on naive method (NM) and JM approach for semiparametric NLME model and PJM associated with nonignorable missing (NIM) or ignorable missing (IM) response data

Missing data	Method	Model	σ_1^2	σ_2^2	δ_e	δ_ϵ	ν_1	ν_2	DIC	EPD	
NIM	JM	I	PM	0.02	0.06	0.24	1.12	3.28	3.79	1368.9	3.85
			L_{CI}	0.01	0.01	0.17	0.84	3.01	3.02		
			U_{CI}	0.03	0.15	0.32	1.33	4.40	5.79		
			SD	0.01	0.04	0.04	0.13	0.28	0.77		
NIM	JM	II	PM	0.09	0.06	0.36	1.75	-	-	572.8	2.41
			L_{CI}	0.05	0.01	0.25	1.47	-	-		
			U_{CI}	0.13	0.25	0.48	1.98	-	-		
			SD	0.02	0.07	0.13	0.13	-	-		
NIM	JM	III	PM	0.13	1.28	-	-	-	-	2599.1	7.83
			L_{CI}	0.10	1.06	-	-	-	-		
			U_{CI}	0.15	1.54	-	-	-	-		
			SD	0.01	0.12	-	-	-	-		
	NM	II	PM	-	0.06	-	1.84	-	-	635.7	3.04
			L_{CI}	-	0.01	-	1.59	-	-		
			U_{CI}	-	0.23	-	2.10	-	-		
			SD	-	0.06	-	0.13	-	-		
NIM	PJM	II	PM	0.07	0.07	0.39	1.79	-	-	1534.5	5.71
			L_{CI}	0.04	0.01	0.22	1.54	-	-		
			U_{CI}	0.11	0.26	0.53	2.03	-	-		
			SD	0.02	0.07	0.08	0.13	-	-		
IM	JM	II	PM	0.09	0.06	0.17	1.80	-	-	1265.7	4.69
			L_{CI}	0.04	0.01	-0.40	1.58	-	-		
			U_{CI}	0.14	0.23	0.54	2.04	-	-		
			SD	0.03	0.06	0.30	0.12	-	-		