The International Journal of Biostatistics

Volume 8, *Issue* 1 2012 *Article* 34

Mixed-Effects Joint Models with Skew-Normal Distribution for HIV Dynamic Response with Missing and Mismeasured Time-Varying Covariate

Yangxin Huang, *University of South Florida*
Jiaqing Chen, *Wuhan University of Technology*
Chunning Yan, *Shanghai University*

Recommended Citation:
Huang, Yangxin; Chen, Jiaqing; and Yan, Chunning (2012) "Mixed-Effects Joint Models with Skew-Normal Distribution for HIV Dynamic Response with Missing and Mismeasured Time-Varying Covariate," *The International Journal of Biostatistics*: Vol. 8: Iss. 1, Article 34.
DOI: 10.1515/1557-4679.1426

©2012 De Gruyter. All rights reserved.

Mixed-Effects Joint Models with Skew-Normal Distribution for HIV Dynamic Response with Missing and Mismeasured Time-Varying Covariate

Yangxin Huang, Jiaqing Chen, and Chunning Yan

Abstract

Longitudinal data arise frequently in medical studies and it is a common practice to analyze such complex data with nonlinear mixed-effects (NLME) models, which enable us to account for between-subject and within-subject variations. To partially explain the variations, time-dependent covariates are usually introduced to these models. Some covariates, however, may be often measured with substantial errors and missing observations. It is often the case that model random error is assumed to be distributed normally, but the normality assumption may not always give robust and reliable results, particularly if the data exhibit skewness. In the literature, there has been considerable interest in accommodating either skewed response or covariate measured with error and missing data in such models, but there has been relatively little study concerning all these features simultaneously. This article is to address simultaneous impact of skewness in response and measurement error and missing data in covariate by jointly modeling the response and covariate processes under a framework of Bayesian semiparametric nonlinear mixed-effects models. In particular, we aim at exploring how mixed-effects joint models based on one-compartment model with one phase time-varying decay rate and two-compartment model with two phase time-varying decay rates contribute to modeling results and inference. The method is illustrated by an AIDS data example to compare potential models with different distributional specifications and various scenarios. The findings from this study suggest that the one-compartment model with a skew-normal distribution may provide more reasonable results if the data exhibit skewness in response and/or have measurement error and missing observations in covariates.

KEYWORDS: Bayesian analysis, HIV dynamics, measurement error, missing covariate, mixed-effects joint models, skew-normal distribution

Author Notes: The authors are extremely grateful to the Editor and two anonymous reviewers for their insightful comments and suggestions that led to a marked improvement of the article. This research was partially supported by NIAID/NIH grant AI080338 to Huang and by the National Natural Science Foundation of China under grant 61174160 to Yan.

1 Introduction

HIV dynamics (by modeling viral load trajectory after initiation of potent antiviral therapy) is one of the most important areas in AIDS research in the last two decades. HIV viral dynamic models have provided new understanding of the pathogenesis of HIV infection and the treatment effects of antiviral therapies. The viral decay rates, as regression parameters of the viral dynamic models, reflect the potency (efficacy) of antiviral therapies (Perelson et al., 1997; Wu and Ding, 1999). A typical feature of the viral load (plasma HIV RNA copies) trajectories is that the inter-individual variation is large (Wu and Ding, 1999, Wu and Wu, 2002a) as shown in Figure 1. It is important to understand the mechanisms of the inter-patient variation in order to make clinical decisions and provide individualized treatments. Much of the inter-individual variation may be explained by some (time-dependent) covariates such as CD4 cell count (the number of $CD4^+$ T lymphocytes per microliter of blood).

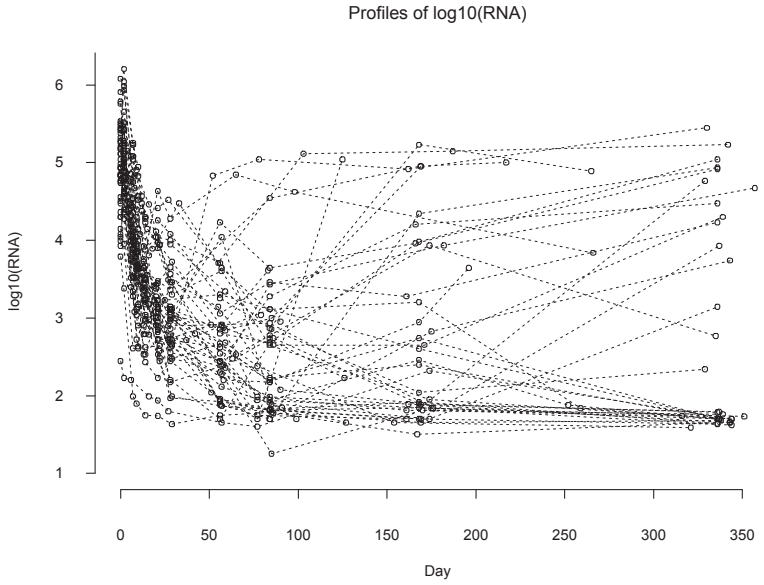

Figure 1: Trajectories of longitudinal viral load measured from RNA levels in plasma (in $\log_{10}$ scale) for 48 patients in an AIDS clinical trial study.

In longitudinal studies, variables, which are called as time-varying (time-dependent), are measured repeatedly over time. Longitudinal data analysis has attracted considerable research interest and, as a result, a large number of statistical modeling and methods have been suggested to analyze such data with

various features in the literature. However, there have been few studies on simultaneously accounting for skewness in responses, measurement error and missing observations in covariates, which are inherent features of longitudinal data. This article proposes statistical joint models and associated inferential methods to address virologic response with asymmetric nature, and measurement errors and missing observations in covariate simultaneously.

Firstly, most of the published methods assume that the error terms in the models for the longitudinal response follow normal distributions due to the mathematical tractability and computational convenience. This requires the variables to be "symmetrically" distributed. A violation of the assumption could lead to misleading inferences. In fact, observed data in AIDS studies are often far from being "symmetric" and asymmetric patterns of observations usually occur. Secondly, measurement error in time-varying covariate (CD4 cell count) is a typical feature of longitudinal data, and ignoring this phenomenon may result in unreasonable statistical inference. Finally, the validity of inference methods relies on an important requirement that variables such as CD4 cell count 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 CD4 time-varying covariate are very common because data may not be always available at each time point, and these missing data may be nonignorable in the sense that the missingness may be related to the values being missing.

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 (Huang and Dagne, 2010; Hughes, 1999; Jara et al., 2008; Wu, 2002, 2004; Wu and Wu, 2002b). However, it is not clear how response asymmetry, covariate missingness and 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 for flexible skew-normal (SN) mixed-effects joint models when these three typical features exist in a longitudinal case study. Under this umbrella, we aim at exploring how mixed-effects joint models based on one-compartment model with one phase time-varying decay rate and two-compartment model with two phase time-varying decay rates (see Appendix A in detail) contribute to modeling results and inference. Specifically, we employed a Bayesian inference approach to jointly investigate the semiparametric nonlinear mixed-effects (SNLME) model with an SN distribution for the viral load response process where viral load trajectories may appear very

complicated patterns, and the linear mixed-effects (LME) model for the CD4 covariate measurement error process where missing observations are present. We consider the SN distribution introduced by Sahu et al.(2003), which is suitable for a Bayesian computation and is briefly discussed in Section 2.2.

The rest of the article is organized as follows. In Section 2, we describe the data set that motivated this research and investigate specific joint models for viral load response dynamics and CD4 covariate measurement error process with missing mechanism. Section 3 presents the associated Bayesian inference method that simultaneously accounts for skewness in response, and missing observations and measurement error in covariate. 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 for HIV dynamics

2.1 Data description

The data set, which motivates this research, is from an AIDS clinical trial study (Lederman et al., 1998). This study consists of 53 HIV-1 infected patients who were treated by an antiretroviral regimen. Five patients who dropped out earlier and never returned to the study were excluded from the data analysis. The plasma HIV-1 RNA (viral load) is repeatedly quantified on days 0, 2, 7, 10, 14, 21, 28, 56, 84, 168 and 336 of follow-up after initiation of treatment. The number of measurements for each individual varies from 7 to 11 . The about 16 percent measurements of viral load are below the limit of detection (LOD), and the LOD of HIV RNA assay is 100 copies per milliliter in this study. For simplicity, we imputed a measurement below LOD by 50, the expected value of those below LOD, which are assumed to take any possible value in [0, 100]. A more formal handling of this issue may be worthwhile, but is beyond the scope of this paper (see more detailed discussion in last Section). CD4 covariate was also measured throughout the study on a similar scheme. The missing rate of the associated CD4 measures was 15% (75 out of 502) at viral load measurement times. The exact day of viral load measurement (not predefined study day) was used to compute study day in our analysis. A $\log_{10}$-transformation of viral load was used in the analysis. Figure 2 shows the measurements of HIV viral load in $\log_{10}$ scale and CD4 cell count for four randomly selected patients. All trajectories of viral load and CD4 cell count exhibit distinctive and important patterns throughout the time course. The rate change in viral load appears to vary substantially across patients,

reflecting both biological variation and systematic associations with subject-level covariates. The detailed descriptions of the study and data can be found in Lederman et al.(1998).

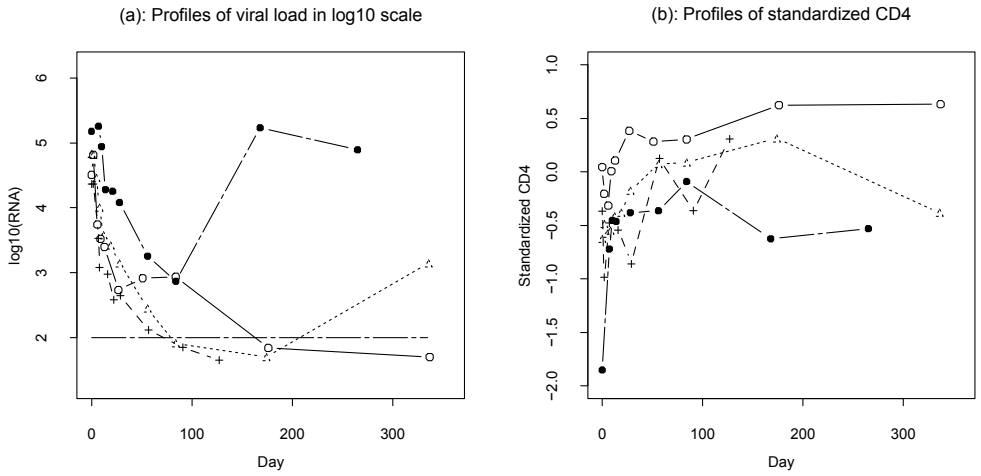

Figure 2: Profiles of viral load (response) in $\log_{10}$ scale and standardized CD4 (covariate) for four randomly selected patients. The horizontal line is below the detectable level of viral load (2=$\log_{10}(100)$).

2.2 Multivariate skew-normal distributions

Recently, there has been an increasing interest in finding more flexible methods to represent features of the data as adequately as possible and to reduce unrealistic assumptions. One approach for data modeling consists in constructing flexible parametric classes of multivariate distributions that is different from the normal distribution. The skew-elliptical distribution is an attractive class of asymmetric thick-tailed parametric structure, which includes the skew-normal (SN) distribution as a special case. Different versions of the multivariate SN distributions have been considered and used in the literature (Arellano-Valle et al., 2005, 2007; Arellano-Valle and Azzalini, 2006; Azzalini and Capitanio, 1999; Jara et al., 2008; Sahu et al., 2003; and others). These studies demonstrated that the SN distribution has reasonable flexibility in real data fitting, while it maintains some convenient formal properties of the normal density. For more detailed discussions on properties and theories of SN distribution and its potential applications as well as differences among various versions of SN distributions, see References listed above.

In this paper, we consider a multivariate SN distribution introduced by Sahu et al.(2003), which is suitable for straightforward Bayesian analysis through hierarchical representations since it is built using conditional method. In particular, it is relatively easy to implement and provides an interesting alternative to other computationally challenging parametric or nonparametric models. For completeness, this section is started by briefly summarizing the multivariate SN distribution that will be used in this paper. An m-dimensional random vector $\boldsymbol{Y}$ follows an m variate SN distribution with location vector $\boldsymbol{\mu}$, $m \times m$ positive (diagonal) dispersion matrix $\boldsymbol{\Sigma}$ and $m \times m$ skewness matrix $\boldsymbol{\Delta} = diag(\delta_1, \delta_2, \ldots, \delta_m)$ with $\boldsymbol{\delta} = (\delta_1, \delta_2, \ldots, \delta_m)^T$ being a skewness parameter vector, if its probability density function (pdf) is given by

$$f(\boldsymbol{y}|\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\Delta}) = 2^m |\boldsymbol{A}|^{-1/2} \phi_m[\boldsymbol{A}^{-1/2}\bar{\boldsymbol{y}}|\boldsymbol{I}_m] \Phi_m[\boldsymbol{\Delta}\boldsymbol{A}^{-1}\bar{\boldsymbol{y}}|\boldsymbol{I}_m - \boldsymbol{\Delta}\boldsymbol{A}^{-1}\boldsymbol{\Delta}], \quad (1)$$

where $\bar{\boldsymbol{y}} = \boldsymbol{y} - \boldsymbol{\mu}$, $\boldsymbol{A} = \boldsymbol{\Sigma} + \boldsymbol{\Delta}^2$, $\phi_m(\boldsymbol{y}|\boldsymbol{V})$ and $\Phi_m(\boldsymbol{y}|\boldsymbol{V})$ denote the pdf and the cumulative distribution function (cdf), respectively, of $N_m(\boldsymbol{0}, \boldsymbol{V})$. We denote this by $\boldsymbol{Y} \sim SN_m(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\Delta})$. The mean and covariance matrix are given by $E(\boldsymbol{Y}) = \boldsymbol{\mu} + \sqrt{2/\pi}\boldsymbol{\delta}$, $cov(\boldsymbol{Y}) = \boldsymbol{\Sigma} + (1 - 2/\pi)\boldsymbol{\Delta}^2$. An appealing feature of the pdf $f(\boldsymbol{y}|\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\Delta})$ is that it gives independent marginal when $\boldsymbol{\Sigma} = diag(\sigma_1^2, \sigma_2^2, \ldots, \sigma_m^2)$. This pdf thus reduces to

$$f(\boldsymbol{y}|\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\Delta}) = \prod_{i=1}^m \left[\frac{2}{\sqrt{\sigma_i^2 + \delta_i^2}} \phi \left\{ \frac{y_i - \mu_i}{\sqrt{\sigma_i^2 + \delta_i^2}} \right\} \Phi \left\{ \frac{\delta_i}{\sigma_i} \frac{y_i - \mu_i}{\sqrt{\sigma_i^2 + \delta_i^2}} \right\} \right], \quad (2)$$

where $\phi(\cdot)$ and $\Phi(\cdot)$ are the pdf and cdf of the standard normal distribution, respectively. It is noted that when $\boldsymbol{\delta} = \boldsymbol{0}$, the SN distribution reduces to usual normal distribution. To better understand the shape of a SN distribution, plots of the univariate SN density as a function of the skewness parameter can be found in publication (Huang and Dagne, 2010).

According to Sahu et al.(2003), if $\boldsymbol{Y}$ follows $SN_m(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\Delta})$, it can be expressed by a convenient stochastic representation as follows.

$$\boldsymbol{Y} = \boldsymbol{\mu} + \boldsymbol{\Delta}|\boldsymbol{X}_0| + \boldsymbol{\Sigma}^{1/2}\boldsymbol{X}_1, \quad (3)$$

where $\boldsymbol{X}_0$ and $\boldsymbol{X}_1$ are two independent random vectors with $N_m(\boldsymbol{0}, \boldsymbol{I}_m)$. Let $\boldsymbol{w} = |\boldsymbol{X}_0|$; then, $\boldsymbol{w}$ follows an m-dimensional standard normal distribution $N_m(\boldsymbol{0}, \boldsymbol{I}_m)$ truncated in the space $\boldsymbol{w} > \boldsymbol{0}$. Thus, a two-level hierarchical representation of (3) is given by

$$\boldsymbol{Y}|\boldsymbol{w} \sim N_m(\boldsymbol{\mu} + \boldsymbol{\Delta}\boldsymbol{w}, \boldsymbol{\Sigma}), \ \boldsymbol{w} \sim N_m(\boldsymbol{0}, \boldsymbol{I}_m)\boldsymbol{I}(\boldsymbol{w} > \boldsymbol{0}). \quad (4)$$

2.3 Skew-normal SNLME joint models for HIV dynamics

As discussed in Appendix A, nonlinear mixed-effects models based on the one-compartment model (A.2) and the two-compartment model (A.3) are powerful tools for modeling HIV viral dynamics (Wu and Ding, 1999) and they offer almost equal performance to capture the early segment of viral load trajectory (Wu and Wu, 2002b). One of our objectives in this paper is to investigate how the two extended equations (A.4) and (A.5) perform when the complete viral load data including viral rebound are employed for modeling. We will develop such model structures and associated inferential methods below.

Denote the number of subjects by n and the number of measurements on the ith subject by n_i. Let y_{ij} be the $\log_{10}$-transformation of the viral load value for individual i at time t_{ij} $(i = 1, \ldots n; j = 1, \ldots, n_i)$. Let $z^*_{ij} = z^*_{ij}(t_{ij})$ be a summary of the true (but unobservable) CD4 covariate value at time t_{ij} (see Section 2.3 in detail). $w(t)$ and $h_i(t)$ are unknown nonparametric smooth fixed-effects and random-effects functions, respectively; $h_i(t)$ are iid realizations of a zero-mean stochastic process. We assume that the vector of model errors $\boldsymbol{e}_i = (e_{i1}, \ldots, e_{in_i})^T \overset{\text{iid}}{\sim} SN_{n_i}\left(-\sqrt{2/\pi}\delta\mathbf{1}_{\mathbf{n_i}}, \sigma_1^2\boldsymbol{I}_{\mathbf{n_i}}, \delta\boldsymbol{I}_{\mathbf{n_i}}\right)$, which follows a multivariate SN distribution with unknown scale parameter σ_1^2, and skewness parameter δ. For the viral load response process, we consider the following two SNLME models with an SN distribution based on equations (A.4) and (A.5), which incorporate possibly mismeasured time-varying CD4 covariate with missing observations.

Model I (one-compartment model with time-varying decay rate):

$$\begin{aligned} y_{ij} &= \log_{10}\left\{e^{p_{i1}} + e^{p_{i2}-\lambda_{ij}(t_{ij})t_{ij}}\right\} + e_{ij}, \\ p_{i1} &= \beta_1 + b_{i1}, \quad p_{i2} = \beta_2 + b_{i2}, \\ \lambda_{ij}(t_{ij}) &= \beta_3 + \beta_4 z^*_{ij} + w(t_{ij}) + h_i(t_{ij}), \end{aligned} \tag{5}$$

where $\exp(p_{i1}) + \exp(p_{i2})$ is the baseline viral load, and λ_{ij} is the viral decay rate; $\boldsymbol{\beta}_{ij} = (p_{i1}, p_{i2}, \lambda_{ij})^T$ is a vector of individual parameters for the i^{th} subject at time t_{ij} and $\boldsymbol{\beta}^\dagger = (\beta_1, \beta_2, \beta_3, \beta_4)^T$ is a vector of population parameters; $\boldsymbol{b}_i^\dagger = (b_{i1}, b_{i2})^T$ is a vector of random-effects.

Model II (two-compartment model with time-varying decay rates):

$$\begin{aligned} y_{ij} &= \log_{10}\left\{e^{p_{i1}-\lambda_{ij1}(t_{ij})t_{ij}} + e^{p_{i2}-\lambda_{ij2}(t_{ij})t_{ij}}\right\} + e_{ij}, \\ p_{i1} &= \beta_1 + b_{i1}, \quad \lambda_{ij1}(t_{ij}) = \beta_3 + \beta_4 z^*_{ij} + b_{i2}, \\ p_{i2} &= \beta_2 + b_{i3}, \quad \lambda_{ij2}(t_{ij}) = w(t_{ij}) + h_i(t_{ij}), \end{aligned} \tag{6}$$

where $\exp(p_{i1}) + \exp(p_{i2})$ is the baseline viral load, and λ_{ij1} and λ_{ij2} are the first- and second phase viral decay rates, respectively; $\boldsymbol{\beta}_{ij} = (p_{i1}, p_{i2}, \lambda_{ij1}, \lambda_{ij2})^T$ is a vector of individual parameters for the i^{th} subject at time t_{ij} and $\boldsymbol{\beta}^\dagger = (\beta_1, \beta_2, \beta_3, \beta_4)^T$ is a vector of population parameters; $\boldsymbol{b}_i^\dagger = (b_{i1}, b_{i2}, b_{i3})^T$ is a vector of random-effects.

Models (5) and (6) accommodate 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. z_{ij}^* and the unknown nonparametric smooth function are incorporated into the decay rates to capture long-term viral load trajectories with different shapes including viral load rebound. These two models are more flexible than commonly-used NLME models. To fit the SNLME models, we apply the regression spline method to nonparametric component. 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 $\boldsymbol{\Psi}_p(t) = \{\psi_0(t), \psi_1(t), ..., \psi_{p-1}(t)\}^T$ and $\boldsymbol{\Phi}_q(t) = \{\phi_0(t), \phi_1(t), ..., \phi_{q-1}(t)\}^T$, respectively. That is,

$$\begin{array}{l} w(t) \approx w_p(t) = \sum_{l=0}^{p-1} \mu_l \psi_l(t) = \boldsymbol{\Psi}_p(t)^T \boldsymbol{\mu}_p, \\ h_i(t) \approx h_{iq}(t) = \sum_{l=0}^{q-1} \xi_{il} \phi_l(t) = \boldsymbol{\Phi}_q(t)^T \boldsymbol{\xi}_{iq}, \end{array} \tag{7}$$

where $\boldsymbol{\mu}_p$ and $\boldsymbol{\xi}_{iq}$ ($q \leq p$ in order 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 $\boldsymbol{\xi}_{iq}$ as iid 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.

Let $\boldsymbol{\beta} = (\boldsymbol{\beta}^{\dagger T}, \boldsymbol{\mu}_p^T)^T$ and $\boldsymbol{b}_i = (\boldsymbol{b}_i^{\dagger T}, \boldsymbol{\xi}_{iq}^T)^T$ follow a multivariate normal distribution $N(\mathbf{0}, \boldsymbol{\Sigma}_b)$, and $\boldsymbol{\Sigma}_b$ is an unstructured covariance matrix. Let $\boldsymbol{y}_i = (y_{i1}, \ldots, y_{in_i})^T$, $\boldsymbol{z}_i^* = (z_{i1}^*, \ldots, z_{in_i}^*)^T$, $\boldsymbol{g}_i(\boldsymbol{t}_i, \boldsymbol{\beta}_i) = (g(t_{i1}, \boldsymbol{\beta}_{i1}), \ldots, g(t_{in_i}, \boldsymbol{\beta}_{in_i}))^T$, $\boldsymbol{t}_i = (t_{i1}, \ldots, t_{in_i})^T$, and $\boldsymbol{\beta}_i = (\boldsymbol{\beta}_{i1}, \ldots, \boldsymbol{\beta}_{in_i})^T$. 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 rewrite Models I and II as follows.

Model I:

$$\begin{array}{l} \boldsymbol{y}_i = \boldsymbol{g}_i(\boldsymbol{t}_i, \boldsymbol{\beta}_i) + \boldsymbol{e}_i, \quad \boldsymbol{e}_i \overset{\text{iid}}{\sim} SN_{n_i}\left(-\sqrt{2/\pi}\delta \mathbf{1}_{\mathbf{n_i}}, \sigma_{\mathbf{1}}^{\mathbf{2}} \boldsymbol{I}_{\mathbf{n_i}}, \delta \boldsymbol{I}_{\mathbf{n_i}}\right), \\ p_{i1} = \beta_1 + b_{i1}, \;\; p_{i2} = \beta_2 + b_{i2}, \\ \lambda_{ij}(t_{ij}) \approx \beta_3 + \beta_4 z_{ij}^* + \boldsymbol{\Psi}_p(t_{ij})^T \boldsymbol{\mu}_p + \boldsymbol{\Phi}_q(t_{ij})^T \boldsymbol{\xi}_{iq}, \end{array} \tag{8}$$

where $g(t_{ij}, \boldsymbol{\beta}_{ij}) = \log_{10}\left\{e^{p_{i1}} + e^{p_{i2} - \lambda_{ij}(t_{ij})t_{ij}}\right\}$.

Model II:

$$\begin{array}{l} \boldsymbol{y}_i = \boldsymbol{g}_i(\boldsymbol{t}_i, \boldsymbol{\beta}_i) + \boldsymbol{e}_i, \quad \boldsymbol{e}_i \overset{\text{iid}}{\sim} SN_{n_i}\left(-\sqrt{2/\pi}\delta \mathbf{1}_{\mathbf{n_i}}, \sigma_{\mathbf{1}}^{\mathbf{2}} \boldsymbol{I}_{\mathbf{n_i}}, \delta \boldsymbol{I}_{\mathbf{n_i}}\right), \\ p_{i1} = \beta_1 + b_{i1}, \;\; \lambda_{ij1}(t_{ij}) = \beta_3 + \beta_4 z_{ij}^* + b_{i3}, \\ p_{i2} = \beta_2 + b_{i2}, \;\; \lambda_{ij2}(t_{ij}) \approx \boldsymbol{\Psi}_p(t_{ij})^T \boldsymbol{\mu}_p + \boldsymbol{\Phi}_q(t_{ij})^T \boldsymbol{\xi}_{iq}, \end{array} \tag{9}$$

where $g(t_{ij}, \boldsymbol{\beta}_{ij}) = \log_{10}\left\{e^{p_{i1} - \lambda_{ij1}(t_{ij})t_{ij}} + e^{p_{i2} - \lambda_{ij2}(t_{ij})t_{ij}}\right\}$. Thus, the SNLME models (8) and (9) revert to parametric NLME models that are similar to those discussed by Huang and Dagne (2010) and Wu (2002). We can express Models I and II in a combined form as follows.

$$\begin{array}{l} \boldsymbol{y}_i = \boldsymbol{g}_i(\boldsymbol{t}_i, \boldsymbol{\beta}_i) + \boldsymbol{e}_i, \quad \boldsymbol{e}_i \overset{\text{iid}}{\sim} SN_{n_i}\left(-\sqrt{2/\pi}\delta \mathbf{1}_{\mathbf{n_i}}, \sigma_{\mathbf{1}}^{\mathbf{2}} \boldsymbol{I}_{\mathbf{n_i}}, \delta \boldsymbol{I}_{\mathbf{n_i}}\right), \\ \boldsymbol{\beta}_i = \boldsymbol{d}(\boldsymbol{z}_i^*, \boldsymbol{\Psi}_p, \boldsymbol{\Phi}_q, \boldsymbol{\beta}, \boldsymbol{b}_i), \quad \boldsymbol{b}_i \overset{\text{iid}}{\sim} N(\mathbf{0}, \boldsymbol{\Sigma}_b), \end{array} \tag{10}$$

2.4 Covariate models with missing and mismeasured data

Various covariate mixed-effects models were investigated in the literature (Carrol et al., 2006; Wu, 2002; and others). This section briefly discusses (CD4) covariate measurement error models with missing observations. Let $z_{ij} = z_{ij}(t_{ij})$ be the observed covariate value for individual i at time t_{ij} $(i = 1, \ldots n; j = 1, \ldots, n_i)$. Some CD4 covariate values may be missing because these covariate values may not be observed at the viral load response measurement time t_{ij}. Let $\boldsymbol{z}_i = (\boldsymbol{z}_{mis,i}, \boldsymbol{z}_{obs,i})$, where $\boldsymbol{z}_{mis,i}$ and $\boldsymbol{z}_{obs,i}$ are the collections of the missing and observed components of $\boldsymbol{z}_i$, respectively. Let $\boldsymbol{r}_i = (r_{i1}, \ldots, r_{in_i})^T$ be a vector of missing CD4 covariate indicator such that $r_{ij} = 1$ if z_{ij} is missing and 0 otherwise.

In the presence of covariate measurement errors, we consider the following LME model to quantify the CD4 covariate process.

$$z_{ij} = \boldsymbol{u}_{ij}^T \boldsymbol{\alpha} + \boldsymbol{v}_{ij}^T \boldsymbol{a}_i + \epsilon_{ij} \;\; (\equiv z_{ij}^* + \epsilon_{ij}), \qquad \boldsymbol{\epsilon}_i \overset{\text{iid}}{\sim} N_{n_i}\left(\mathbf{0}, \sigma_2^2 \boldsymbol{I}_{n_i}\right), \tag{11}$$

where $z_{ij}^* = z_{ij}^*(t_{ij}) = \boldsymbol{u}_{ij}^T \boldsymbol{\alpha} + \boldsymbol{v}_{ij}^T \boldsymbol{a}_i$ may be viewed as the true (but unobservable) CD4 covariate value at time t_{ij}, $\boldsymbol{u}_{ij} = \boldsymbol{u}_{ij}(t_{ij})$ and $\boldsymbol{v}_{ij} = \boldsymbol{v}_{ij}(t_{ij})$ are

$l \times 1$ design vectors, $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_l)^T$ and $\boldsymbol{a}_i = (a_{i1}, \ldots, a_{il})^T$ are unknown population (fixed-effects) and individual-specific (random-effects) parameter vectors, respectively. The random-effects $\boldsymbol{a}_i$, which are introduced to account for large inter-individual variations in the CD4 process, follow the multivariate normal distribution $N(0, \boldsymbol{\Sigma}_a)$, where $\boldsymbol{\Sigma}_a$ is unrestricted covariance matrix. We assume $\boldsymbol{e}_i, \boldsymbol{b}_i, \boldsymbol{\epsilon}_i$ and $\boldsymbol{a}_i$ are independent of each other.

To allow for nonignorable missing mechanism in the CD4 covariate, we should assume a missing data model for the missing covariate mechanism. Thus, we focus on the following simple independent missing data model (to avoid too many nuisance parameters) although more complicated missing covariate models can also be considered.

$$f(\boldsymbol{r}|\boldsymbol{\eta}) = \prod_{i=1}^{n} f(\boldsymbol{r}_i|\boldsymbol{\eta}) = \prod_{i=1}^{n}\prod_{j=1}^{n_i} [P(r_{ij} = 1|\boldsymbol{\eta})]^{r_{ij}} [1 - P(r_{ij} = 1|\boldsymbol{\eta})]^{1-r_{ij}}, \quad (12)$$

where $\text{logit}[P(r_{ij} = 1|\boldsymbol{\eta})] = \eta_0 + \eta_1 z_{ij}$, where $\boldsymbol{\eta} = (\eta_0, \eta_1)^T$ is a vector of unknown nuisance parameters.

As pointed out previously, the main goal of this article is to compare performance on how Model I (based on one-compartment model with one phase time-varying decay rate) and Model II (based on two-compartment model with two phase time-varying decay rates) contribute to modeling results and parameter estimation. It is noticed that Model II discussed in this article is equivalent to that studied by Huang and Dagne (2012), but the different data features were considered. In addition, the parametric NLME models investigated by Huang and Dagne (2010, 2011) were special cases of Model II presented in this article in the sense that Model II reduces to the models conducted by Huang and Dagne (2010, 2011) when the following further treatments are made. (i) The unknown nonparametric smooth function in the model (9) is reverted to known explicit function, i.e., $w(t_{ij}) = \beta_5 + \beta_6 x_{ij} + \beta_7 z_{ij}$, where x_{ij} is CD8 values (Huang and Dagne, 2010) or $w(t_{ij}) = \beta_5 + \beta_6 z^*_{ij}$ (Huang and Dagne, 2011) and $h(t_{ij}) = b_{i4}$. (ii) The missing covariate data models are not considered in those models. (iii) The true (but unobserved) time-varying CD4 values z^*_{ij} are replaced by baseline CD4 value in the first-phase viral decay rate λ_{ij1}.

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 distribution 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 response model (10) and the covariate model (11) associated with the missing data model to estimate all parameters simultaneously, and we obtain numerical approximations to posterior distributions using MCMC procedure.

Following discussion in Sahu et al.(2003), in order to implement an MCMC procedure to the joint model, it can be shown by introducing an $n_i \times 1$ random vector $\boldsymbol{w}_{e_i}$ based on the stochastic representation for the SN distribution (see Section 2.2 in detail) that $\boldsymbol{y}_i$ and $\boldsymbol{z}_i$ associated with the missing data model (12) can be hierarchically formulated as follows.

$$\begin{array}{l} \boldsymbol{y}_i|\boldsymbol{z}_i, \boldsymbol{a}_i, \boldsymbol{b}_i, \boldsymbol{w}_{e_i}; \boldsymbol{\alpha}, \boldsymbol{\beta}, \sigma_1^2, \delta \sim N_{n_i}\left(\boldsymbol{g}_i(\boldsymbol{t}_i, \boldsymbol{\beta}_i) + \delta \boldsymbol{w}_{e_i}, \sigma_1^2 \boldsymbol{I}_{n_i}\right), \\ \boldsymbol{z}_i|\boldsymbol{a}_i, \boldsymbol{r}_i; \boldsymbol{\alpha}, \sigma_2^2 \sim N_{n_i}\left(\boldsymbol{z}_i^*, \sigma_2^2 \boldsymbol{I}_{n_i}\right),\ r_{ij}|\boldsymbol{\eta} \sim Bernoulli(P_{ij}), \\ \boldsymbol{w}_{e_i} \sim N_{n_i}(\boldsymbol{0}, \boldsymbol{I}_{n_i}) I(\boldsymbol{w}_{e_i} > \boldsymbol{0}),\ \boldsymbol{b}_i \sim N(\boldsymbol{0}, \boldsymbol{\Sigma}_b),\ \boldsymbol{a}_i \sim N(\boldsymbol{0}, \boldsymbol{\Sigma}_a), \end{array} \tag{13}$$

where $I(\boldsymbol{w} > 0)$ is an indicator function and $\boldsymbol{w} = |\boldsymbol{\varsigma}|$ with $\boldsymbol{\varsigma} \sim N_{n_i}(\boldsymbol{0}, \boldsymbol{I}_{n_i})$, and $P_{ij} = P(r_{ij} = 1|\boldsymbol{\eta})$. 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 of the model with an SN distribution is almost equivalent to that with a normal distribution.

Let $\boldsymbol{\theta} = \{\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\eta}, \sigma_1^2, \sigma_2^2, \boldsymbol{\Sigma}_a, \boldsymbol{\Sigma}_b, \delta\}$ be the collection of unknown population parameters in models (10), (11) and (12). Under the Bayesian framework, we next need to specify prior distributions for all the unknown parameters in these models as follows.

$$\begin{array}{l} \boldsymbol{\beta} \sim N(\boldsymbol{\tau}_1, \boldsymbol{\Lambda}_1),\ \sigma_1^2 \sim IG(\omega_1, \omega_2),\ \boldsymbol{\Sigma}_b \sim IW(\boldsymbol{\Omega}_1, \rho_1),\ \delta \sim N(0, \gamma), \\ \boldsymbol{\alpha} \sim N(\boldsymbol{\tau}_2, \boldsymbol{\Lambda}_2),\ \sigma_2^2 \sim IG(\omega_3, \omega_4),\ \boldsymbol{\Sigma}_a \sim IW(\boldsymbol{\Omega}_2, \rho_2),\ \boldsymbol{\eta} \sim N(\boldsymbol{\tau}_3, \boldsymbol{\Lambda}_3), \end{array} \tag{14}$$

where the mutually independent Normal (N), Inverse Gamma (IG), and Inverse Wishart (IW) prior distributions are chosen to facilitate computations. The super-parameter matrices $\boldsymbol{\Lambda}_1$, $\boldsymbol{\Lambda}_2$, $\boldsymbol{\Lambda}_3$, $\boldsymbol{\Omega}_1$ and $\boldsymbol{\Omega}_2$ can be assumed to be diagonal for convenient implementation.

Let the observed data $\mathcal{D} = \{(\boldsymbol{y}_i, \boldsymbol{z}_{obs,i}, \boldsymbol{r}_i), i = 1, \ldots, n\}$. 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 $\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\eta}, \sigma_1^2, \sigma_2^2, \boldsymbol{\Sigma}_a, \boldsymbol{\Sigma}_b$ and δ are independent of each other, i.e., $\pi(\boldsymbol{\theta}) = \pi(\boldsymbol{\alpha})\pi(\boldsymbol{\beta})\pi(\boldsymbol{\eta})\pi(\sigma_1^2)\ \pi(\sigma_2^2)\pi(\boldsymbol{\Sigma}_a)$ $\pi(\boldsymbol{\Sigma}_b)\pi(\delta)$. 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 covariate data 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

$$\begin{aligned} f(\boldsymbol{\theta}|\mathcal{D}) \quad \propto \quad & \{\prod_{i}^{n} \int\int\int f(\boldsymbol{y}_i|\boldsymbol{z}_i, \boldsymbol{a}_i, \boldsymbol{b}_i, \boldsymbol{w}_{e_i}; \boldsymbol{\alpha}, \boldsymbol{\beta}, \sigma_1^2, \delta) f(\boldsymbol{z}_i|\boldsymbol{a}_i, \boldsymbol{r}_i; \boldsymbol{\alpha}, \sigma_2^2) \\ & f(\boldsymbol{w}_{e_i}|\boldsymbol{w}_{e_i} > \mathbf{0}) f(\boldsymbol{r}_i|\boldsymbol{\eta}) f(\boldsymbol{a}_i) f(\boldsymbol{b}_i) d\boldsymbol{z}_{mis,i} d\boldsymbol{a}_i d\boldsymbol{b}_i\} \pi(\boldsymbol{\theta}). \end{aligned} \tag{15}$$

In general, the integrals in (15) 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 (15), using the Gibbs sampler along with the Metropolis-Hastings (M-H) algorithm.

4 Analysis of AIDS clinical data

4.1 Model implementation

Models (8) and (9) in Section 2.2, which specify the viral decay rate nonparametrically, appear 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 (2006) for likelihood inference, we set $\psi_0(t) = \phi_0(t) \equiv 1$ and take the same natural cubic splines in the approximations (7) with $q \leq p$. The values of p and q are determined by the AIC/BIC criteria. The AIC/BIC values are evaluated by various (p, q) combinations $(p, q) = \{(1,1), (2,1), (2,2), (3,1), (3,2), (3,3)\}$, which suggest the following function for $w(t_{ij})$ with $p = 3$ and $h_i(t_{ij})$ with $q = 1$:

$$w(t_{ij}) + h_i(t_{ij}) \approx \beta_5 + \beta_6 \psi_1(t_{ij}) + \beta_7 \psi_2(t_{ij}) + \xi_{i0}. \tag{16}$$

Thus, we have population parameter vector $\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_7)^T$ and individual random-effects $\boldsymbol{b}_i = (b_{i1}, b_{i2}, \xi_{i0})^T$ in Model I or $\boldsymbol{b}_i = (b_{i1}, b_{i2}, b_{i3}, \xi_{i0})^T$ in Model II follows $N(\mathbf{0}, \boldsymbol{\Sigma}_b)$.

Figure 2(b) shows the CD4 trajectories of four randomly-selected patients. It is known that CD4 is often measured with substantial error, so

it is reasonable to assume that time-varying decay rate in model (10) 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 and missing data in the CD4 covariate. 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 (11) with $\boldsymbol{u}_{ij} = \boldsymbol{v}_{ij} = (1, t_{ij}, \ldots, t_{ij}^{l-1})^T$ and focus on linear ($l = 2$), quadratic ($l = 3$) and cubic ($l = 4$) polynomials. The resulting AIC (BIC) values are 890.0 (931.7), 773.4 (801.3) and 845.2 (887.1), respectively. Thus, we adopt the following quadratic polynomial LME model for the CD4 process:

$$z_{ij} = (\alpha_1 + a_{i1}) + (\alpha_2 + a_{i2})t_{ij} + (\alpha_3 + a_{i3})t_{ij}^2 + \epsilon_{ij}, \tag{17}$$

where $z_{ij}^* = (\alpha_1 + a_{i1}) + (\alpha_2 + a_{i2})t_{ij} + (\alpha_3 + a_{i3})t_{ij}^2$, $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \alpha_3)^T$ is a vector of population (fixed-effects) parameters and individual-specific random-effects $\boldsymbol{a}_i = (a_{i1}, a_{i2}, a_{i3})^T \sim N(\mathbf{0}, \boldsymbol{\Sigma}_a)$, and $\boldsymbol{\epsilon}_i = (\epsilon_{i1}, \ldots, \epsilon_{in_i})^T \sim N_{n_i}(\mathbf{0}, \sigma_2^2 \boldsymbol{I}_{n_i})$. In addition, in order to avoid too small or large estimates that may be unstable, we standardize the time-varying covariate CD4 cell count (each CD4 value is subtracted by 258.6 and divided by 100.3) and rescale the original time t (in days) so that the time scale is between 0 and 1.

We investigated the following four scenarios. First, for modeling HIV viral dynamics with constant decay rates, the one-compartment model (A.2) and two-compartment model (A.3) perform similarly at earlier treatment period (Wu and Wu, 2002b). In other words, the models were used to fit truncated data only where individual viral load trajectory shows a decrease pattern. We investigated how Models I and II with time-varying decay rates contribute to modeling results and parameter estimation based on complete data at whole study period. Second, since a normal distribution is a special case of an SN distribution when the skewness parameter is zero, we investigated how an asymmetric (SN) distribution for model error contributes to modeling results and parameter estimation in comparison with a symmetric (normal) distribution. Third, we estimated the model parameters by using the "naive" (denoted by NV) method, which ignores measurement errors in CD4 covariate. That is, the NV method uses the observed CD4 values z_{ij} rather than unobservable CD4 values z_{ij}^* in the response model (10). We used it as a comparison to the joint modeling (JM) approach proposed in this paper. This comparison attempted to investigate how the measurement errors in CD4 covariate contribute to modeling results. Finally, when covariates are measured

with errors, a common approach is the so-called two-step (TS) method (Higgins et al., 1997): the first step estimates the 'true' covariate value based on the covariate model (17); at the second step the covariate z_{ij}^* in the response model (10) is substituted by its estimate from the first step. Thus, the two-step (TS) method is provided to compare the performance with the JM method.

To carry out the Bayesian inference, we specified the values of the hyper-parameters in the prior distributions of the population parameters. We took weakly-informative prior distributions for the parameters in the joint model. In particular, (i) fixed-effects were taken to be independent normal distribution $N(0, 100)$ for each component of the population parameter vectors $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ and $\boldsymbol{\eta}$; (ii) for the scale parameters σ_1^2 and σ_2^2 we assumed a non-informative 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 $\boldsymbol{\Sigma}_a$ and $\boldsymbol{\Sigma}_b$ were taken to be inverse Wishart distributions $IW(\boldsymbol{\Omega}_1, \rho_1)$ and $IW(\boldsymbol{\Omega}_2, \rho_2)$, where the diagonal elements for diagonal covariance matrices $\boldsymbol{\Omega}_1$ and $\boldsymbol{\Omega}_2$ were 0.01, and $\rho_1 = \rho_2 = 4$; and (v) for the skewness parameter δ, we chose normal distribution $N(0, 100)$.

The MCMC sampler was implemented using WinBUGS package (Lunn et al., 2000), and the WinBUGS program code is available in Appendix B. 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 $\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\eta}, \sigma_1^2, \sigma_2^2, \boldsymbol{\Sigma}_a$, $\boldsymbol{\Sigma}_b$, and δ; (ii) we update $\boldsymbol{b}_i$ and $\boldsymbol{a}_i$ $(i = 1, 2, \cdots, n)$ using the Metroplis-Hastings (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 et al.(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 hyper-parameters, 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 package(such as trace plots). After convergence was achieved, one long chain was run with the following considerations. We proposed that, after an initial number of 50,000 burn-in iterations, every 20th MCMC sample was retained from the next 200,000. Thus, we obtained 10,000 samples of targeted posterior distributions of the unknown parameters for statistical inference.

The Bayesian joint modeling approach based on the SNLME models (10) in conjunction with the approximated nonparamteric component (16) and the covariate measurement error model (17) associated with missing data

model (12) was used to fit the viral load data and CD4 data with measurement error and missing observations. In the following sections, we will report analysis results of the four scenarios proposed in Section 4.1.

Table 1: Summary of estimated posterior means (PM) for population parameters of fixed-effects, precision and skewness, and corresponding standard deviation (SD) and lower limit (L_{CI}) and upper limit (U_{CI}) of 95% equal-tail credible intervals (CI) based on the joint modeling(JM), naive (NV) and two-step (TS) methods with skew-normal (SN) or normal (N) distributions.

Method	Model		α_1	α_2	α_2	β_1	β_2	β_3	β_4	σ_1^2	σ_2^2	δ
JM	I-SN	PM	-0.34	3.68	-3.40	6.03	11.3	36.8	-10.6	0.01	0.31	0.91
		L_{CI}	-0.56	2.83	-4.27	5.57	10.7	22.1	-14.7	0.00	0.26	0.83
		U_{CI}	-0.13	4.48	-2.54	6.52	11.8	51.9	-2.22	0.04	0.36	0.99
		SD	0.11	0.42	0.44	0.23	0.27	7.61	6.00	0.01	0.03	0.04
	II-SN	PM	-0.35	3.76	-3.48	11.1	5.84	69.8	-14.2	0.05	0.31	0.63
		L_{CI}	-0.58	2.98	-4.30	10.6	5.15	61.9	-22.7	0.02	0.26	0.50
		U_{CI}	-0.13	4.54	-2.65	11.6	6.53	78.0	-4.52	0.10	0.36	0.75
		SD	0.11	0.39	0.42	0.26	0.34	4.08	4.68	0.02	0.03	0.06
	I-N	PM	-0.35	3.67	-3.39	5.22	9.97	34.1	-8.17	0.31	0.31	–
		L_{CI}	-0.56	2.90	-4.18	4.69	8.50	22.1	-17.7	0.27	0.26	–
		U_{CI}	-0.12	4.45	-2.58	6.74	11.5	51.7	1.60	0.36	0.36	–
		SD	0.11	0.40	0.40	0.27	0.25	7.52	5.98	0.02	0.03	–
NV	I-SN	PM	-0.33	3.62	-3.36	6.02	11.3	40.0	-7.24	0.01	0.30	0.92
		L_{CI}	-0.54	2.83	-4.16	5.61	10.7	24.3	-13.7	0.00	0.26	0.84
		U_{CI}	-0.11	4.40	-2.53	6.42	11.8	53.6	-0.24	0.03	0.36	1.00
		SD	0.11	0.40	0.42	0.21	0.29	7.21	3.46	0.01	0.02	0.04
TS	I-SN	PM	-0.34	3.63	-3.36	6.04	11.3	37.7	-9.94	0.01	0.30	0.91
		L_{CI}	-0.56	2.86	-4.16	5.63	10.7	22.3	-20.2	0.00	0.26	0.83
		U_{CI}	-0.12	4.41	-2.55	6.40	11.8	53.3	-1.42	0.04	0.36	0.99
		SD	0.11	0.40	0.41	0.20	0.25	6.98	5.53	0.01	0.03	0.04

4.2 Comparison of results between Models I and II

The posterior mean (PM), the corresponding standard deviation (SD) and 95% credible interval for fixed-effects parameters based on the three methods (JM, NV and TS methods) are presented in Table 1. The following findings are observed for estimated results of parameters in Models I and II based on JM method. (i) In the response model, the findings for the most interested parameters (β_3, β_4), which are relative to viral decay rate, show that these estimates are statistically significant for both Models I and II since the 95%

credible intervals do not contain zero. Nevertheless, for the estimate of the coefficient of CD4 covariate β_4, it is shown that CD4 covariate has a significantly negative effect on the viral decay rate, suggesting that the CD4 covariate may be an important predictor of the viral decay rate during the treatment. For the precision parameter σ_1^2, the estimated value (0.01) based on Model I is much smaller than that (0.05) based on Model II. The estimates of the skewness parameter (δ) of Models I and II are, respectively, 0.91 with 95% credible intervals $(0.83, 0.99)$ and 0.63 with 95% credible intervals $(0.50, 0.75)$. This finding suggests that there is a significantly positive skewness in the viral load data and confirms the fact that the distribution of the original data is skewed even after taking ln-transformation. Thus, incorporating a skewness parameter in the modeling of the data is recommended. (ii) For parameter estimates of the CD4 covariate model (17), the estimates of the coefficients based on the two models are comparable but the estimated α_2 is significantly positive, whereas the estimates of α_1 and α_3 are significantly negative. This finding suggests that there is a positive linear relation between CD4 cell count and measurement time. There is an identical estimate (0.31) in the posterior mean of the scale parameter σ_2^2 in Models I and II.

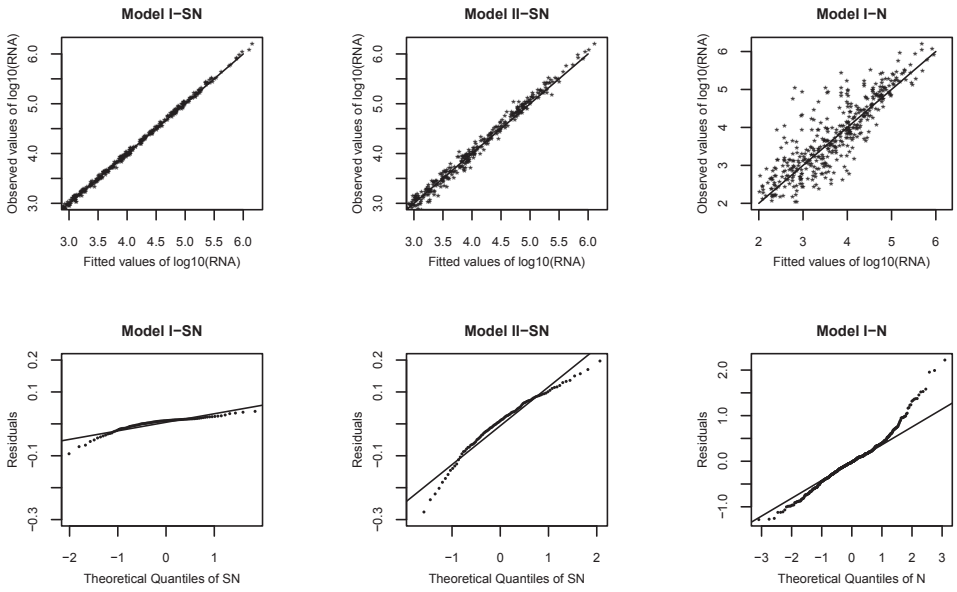

Figure 3: Goodness-of-fit. (1) Top panel: Observed values versus fitted values of $\log_{10}$(RNA); (2) Bottom panel: SN and N (normal) Q-Q plots with line.

Note that the nonignorable model (12) is not testable based on the observed data, so it is important to carry out sensitivity analyses based on

alternative missing data models. For example, we can consider the following alternative missing data models: $\text{logit}[P(r_{ij}=1|\boldsymbol{\eta})] = \eta_0 + \eta_1 z_{i,j-1} + \eta_2 z_{ij}$ and $\text{logit}[P(r_{ij}=1|\boldsymbol{\eta})] = \eta_0 + \eta_1 z_{ij}^2$. The resulting estimates are similar (results are not present here), so the results reported in Table 1 are robust against missing data models.

From the model fitting results, we have seen that, in general, both Model I and Model II provided a reasonably good fit to the observed data for most patients in our study, although the fitting for a few patients (<7%) was not completely satisfactory due to unusual viral load fluctuation patterns for these patients. To assess the goodness-of-fit of the proposed models, the diagnosis plots of the observed values versus the fitted values (top panel) and SN or normal Q-Q plots (bottom panel) from Models I and II are presented in Figure 3. It can be seen from Figure 3 (top panel) that Model I with SN distribution provided better fit to observed data, compared with Model II with SN distribution. This result can be also explained by examining the SN Q-Q plots of the residuals (bottom panel) that both plots show the existence of outliers, but it is clearly seen that Model I only has few negative outliers, and thus, fit observed data better than Model II. This finding is further confirmed by their residual sums of squares (RSS) summarized in Table 2, which are 6.315 (Model I) and 25.44 (Model II).

Table 2: Summary of DIC value, expected predictive deviance (EPD) and residual sum of squares (RSS) based on the joint modeling(JM), naive (NV) and two-step (TS) methods with skew-normal (SN) or normal (N) distributions.

Method	Model	DIC	EPD	RSS
JM	I-SN	1571.8	0.025	6.315
	II-SN	2189.1	0.102	25.44
	I-N	2113.9	0.625	156.5
NV	I-SN	1635.7	0.044	10.68
TS	I-SN	1587.9	0.031	6.394

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 an alternative, we also evaluate expected predictive deviance (EPD) formulated by $EPD = E\left\{\sum_{i,j}(y_{rep,ij} - y_{obs,ij})^2\right\}$ for model 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 $\boldsymbol{\theta}$ (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 joint modeling approach based on Models I and II, which are 1571.8 and 2189.1, respectively. As mentioned previously, it is difficult to tell which model is "correct", only which one fits the data better. Therefore, based on the DIC, the results indicate that Model I provides better fit than Model II. This finding is confirmed by the results of the EPD values (see Table 2). These results are also consistent with those in diagnosis of the goodness-of-fit displayed in Figure 3, indicating that Model I performs better. Based on these observations, we will further report our results in detail only for the better Model I below.

4.3 Results based on Model I

We compare the inferential results between Model I with SN distribution (SN model) and Model I with normal (N) distribution (N model). To assess the goodness-of-fit, the diagnosis plots of the observed values versus the fitted values (top panel), and SN and N Q-Q plots (bottom panel) are presented in Figure 3. It can be seen from Figure 3 (top panel) that the SN model provided better fit to observed data, compared with the N model. This result can be also explained by examining the SN or N Q-Q plots of the residuals (bottom panel) that both plots show the existence of outliers, but it is clearly seen that the SN model only has few negative outliers, and thus, fit observed data better than the N model. This finding is consistently confirmed by the three model selection criteria (DIC, EPD and RSS).

Figure 4 displays estimates of viral load and standardized CD4 trajectories obtained by using the JM method based on the SN and N models for two of the four individual subjects shown in Figure 2. The following findings are observed from joint modeling results. For viral load response (top panel), (i) the estimated individual trajectories for the SN model fit the originally observed viral load values more closely than those for the N model. Note that the lack of smoothness for estimates of individual trajectories in the SN model is understandable since a random component $\boldsymbol{w}_{e_i}$ was incorporated in the expected function (see (13) for details) according to the stochastic representation feature of the SN distribution for "chasing the data" to this extent; (ii) overall, the 95% credible interval (CI) (the last five CIs are shown in the plot for each individual subject) associated with predicted value from the N

model is wider than the corresponding 95% CI from the SN model; (iii) all the 95% CIs from the SN model cover the true (observed) viral load values, while some of 95% CIs from the N model do not. For example, for patient 31, the observed value at day 336 is 3.146; the corresponding 95% CI from Model I with SN distribution is (2.881,3.341) with the fitted value 3.121, but the corresponding 95% CI from Model I with N distribution is (1.988,3.014) with the fitted value 2.753. For CD4 covariate, the plots (bottom panel) indicate that the quadratic model, which is the best polynomial model based on AIC/BIC values, may be reasonable for most individual CD4 trajectories. Note that fitting curves from both the SN and N models are very similar due to the fact that, in this study, the SN distribution reflects to the response models only and the normal distribution is assumed for the CD4 covariate model (17).

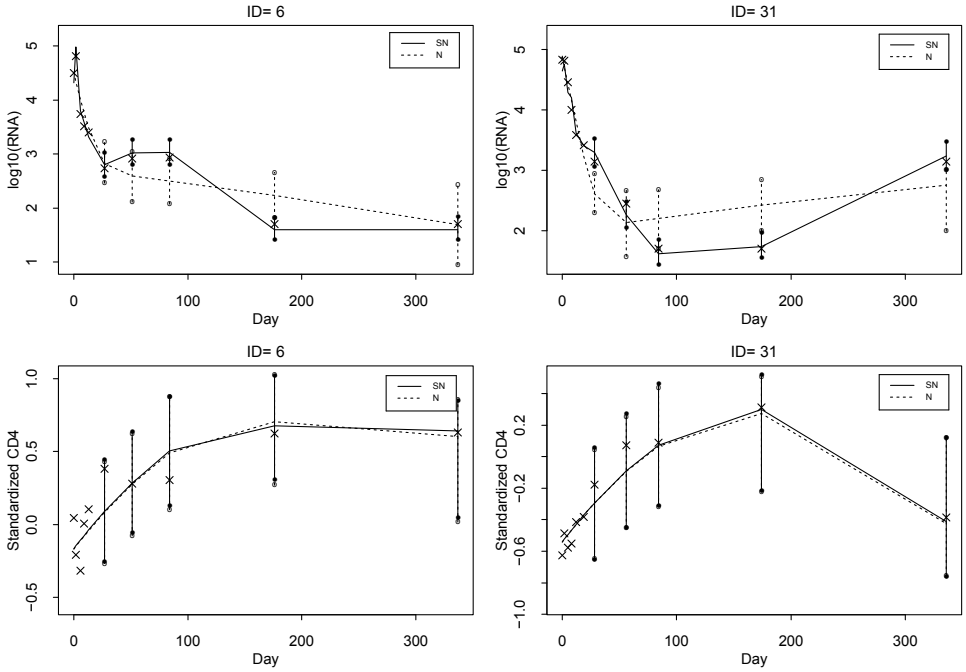

Figure 4: The individual fitted curves of viral load and CD4 for the two representative patients based on the joint models with the normal distribution, denoted by N (dotted line) or SN distribution (solid line) for response model error. The respective vertical dotted line (N) ended with '$\circ$' and solid line (SN) ended with '$\bullet$' on each fitted value are the 95% credible interval (CI) associated with the fitted value. The observed values are indicated by sign crosses ($\times$).

For parameter estimates of the response model, the estimates of β_1, β_2 and β_3 based on the SN model are significantly larger than those based

on the N model, while the estimate of β_4 based on the SN model, which is coefficient of true CD4 covariate, is significantly smaller than that based on the N model. The results indicate that estimated parameters may be substantially underestimated or overestimated if model distribution ignores skewness. For the precision parameter σ_1^2, the estimated value (0.01) based on the SN model is much smaller than that (0.31) based on the N model. (ii) For parameter estimates of the CD4 covariate model, all of the estimated parameters are almost identical for both the SN and N models. (iii) For the missing data model, the estimates of η_0 and η_1 are -1.93 and -0.12, respectively, for both the SN and N models. The results indicate that the chance of missing data is decreasing with CD4 counts.

To investigate how the measurement errors with nonignorable missing observations in CD4 covariate contribute to modeling results, we used the 'naive' (NV) method based on Model I, where the true (unobserved) CD4 values z_{ij}^* associated with the viral decay rate λ_{ij} are substituted by the raw (observed) CD4 values z_{ij} in the response model (8), to estimate the model parameters. It can be seen from Table 1 that there are important differences in the estimates of the parameters β_3 and β_4, which are directly associated with whether or not ignoring potential CD4 measurement errors with missing mechanism for inference. The NV method may substantially overestimate the covariate CD4 effect (β_4). The estimated standard deviation for the CD4 effect (β_4) using the JM method is almost twice larger than that using the NV method. The difference of the naive estimates and the JM estimates, due to whether or not ignoring potential CD4 measurement errors with missing data, indicates that CD4 measurement errors can not be ignored in the analysis. We also obtain estimated DIC, EPD and RSS values (see Table 1) based on the NV method, and the results show that the JM approach provides a better fit to the data in comparison with the NV method.

Table 1 also presents estimated population parameters using the two-step (TS) method. We can see the TS estimates and the JM estimates are somewhat different. There are slight differences in the estimates for the parameters β_3 and β_4, which are directly associated CD4 covariate. For the parameters β_k ($k = 1, 2, 3, 4$), the standard deviations (SD) from the TS method is smaller than those obtained by the JM method. This is because the usual TS method ignores the variability due to estimating the parameters in the covariate model. So the SD produced by the TS method may be underestimated.

The estimated results based on JM approach (Model I) indicate that the population CD4 trajectory may be approximated by the quadratic polynomial $\hat{z}(t) = 100.3(-0.34 + 3.68t - 3.40t^2) + 258.6$, where $\hat{z}(t)$ is in the original CD4 scale. The estimated population viral decay rate is $\hat{\lambda}(t) = 36.8 - 10.6\hat{z}(t) +$

$37.0 - 1.80\psi_1(t) - 0.24\psi_2(t)$. Thus, the population viral load process may be approximated by $\hat{V}(t) = \exp(6.03) + \exp(11.3 - \hat{\lambda}(t)t)$. Since the viral decay rate $\lambda(t)$ is significantly associated with the true CD4 value (due to the statistically significant estimate of β_4), this suggests that the viral load change $V(t)$ may be significantly associated with the CD4 covariate. 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.

The estimated results based on the JM approach (Model I) in Table 1 show that the estimate of skewness parameter (0.91) in viral load is significantly positive, which confirms the right tail skewness of the viral load. Thus, it may suggest that accounting for significant skewness, when the data exhibit asymmetry, provides a better model fit to the data and gives more reasonable estimates to the parameters.

5 Discussion

This paper developed a joint modeling approach for the SNLME model with an SN distribution under a Bayesian framework for longitudinal data with skewness characteristics of viral response and measurement errors and missing data in CD4 covariate, that may be preferred over that with a standard normal distribution. The foregoing results indicate that the one-compartment HIV dynamic model performed better than the two-compartment HIV dynamic model. The results also indicate that the JM method outperformed the NV and TS methods in the sense that the JM method may produce more robust and reasonable parameter estimates. The proposed JM method is quite general, and so can be used to other applications. This kind of SN modeling approach is important in many biostatistical application areas, allowing accurate inference of parameters while adjusting for the data with skewness. The SN distribution is shown to provide an alternative to (symmetric) normal distribution that is often assumed in statistical models.

The proposed SNLME joint model with SN distribution can be easily fitted via MCMC procedure using the publicly-available WinBUGS package and has a computational cost similar to the normal version of the model due to the features of its hierarchically stochastic representation. Implementation via MCMC sampler makes it straightforward to compare the proposed models and methods with various scenarios for real data analysis in comparison with symmetric distributions and asymmetric distributions for model errors. This

makes our approach quite powerful and also accessible to practicing statisticians in the fields. It is noted that Huang and Dagne (2010; 2011) investigated the NLME model with assuming both the model error and random-effects to have the SN distribution and found that the modeling results based on the SN distribution for random-effects are similar to those based on a normal distribution for random-effects. Thus, the random-effects were assumed to be normal here. In addition, the CD4 covariate models were assumed to follow normal distribution since the CD4 cell count data appeared to be normally distributed in this studies although more robust distributional specification such as SN distribution can be employed.

In order to examine the sensitivity of parameter estimates to the prior distributions and initial values, we also conducted a limited sensitivity analysis using different values of hyper-parameters of prior distributions and different initial values (data not shown). The results of the sensitivity analysis showed that the estimated dynamic parameters were not sensitive to changes of both priors and initial values. Thus, the final results are reasonable and robust, and the conclusions of our analysis remain unchanged (see Huang et al. (2006) for more details).

The unreliable observations (actually observed values) below LOD may be used instead of such values imputed by half of LOD or more advanced statistical methods may be conducted to evaluate values below LOD for inference, but results of parameter estimates may be interpreted differently. For example, one of approaches to handle left-censoring due to LOD was adopted by Wu (2002) in which the observed value y_{ij} is denoted by $(q_{ij}; c_{ij})$, where c_{ij} is the censoring indicator such that y_{ij} is observed if $c_{ij} = 0$ and y_{ij} is left censored if $c_{ij} = 1$, i.e. $y_{ij} = q_{ij}$ if $c_{ij} = 0$, and $y_{ij} \leq d$ if $c_{ij} = 1$, where $d = 100$ (e.g. LOD). Thus, the values of LOD can be predicted using Bayesian modeling approach investigated in this paper.

It is noted that there is the possibility of interactively-jointly modeling CD4 counts and viral load based on the HIV dynamic models formulated through a system of ordinary differential equations (ODE)(Guedj et al., 2007; Huang et al., 2006; Wu and Ding, 1999), rather than treating CD4 counts as explanatory variable. As discussed in Appendix A, the proposed SNLME joint models (10) are constructed based on equations (A.4) and (A.5), which are approximated from biologically-based viral dynamic ODE model (A.1). It has seen that the SNLME joint models not only perform well, but also offer the mathematical tractability and computational efficiency. We do not directly adopt biologically-based viral dynamic ODE model (A.1) with the following consideration. Although such model can provide more meaningful biological interpretation for the results, it may involve intensive computational burden

and cause the difficulty of ODE numerical solution, in particular, when joint models-based ODE systems for longitudinal data with features of skewness, and missing and mismeasured covariate are considered.

This article considers two statistical models–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 in the presence of 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 distribution 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.

In summary, for this particular data set based on Model I, which is preferable to Model II, our data analysis indicates that for reliable estimation of HIV dynamic parameters, we should simultaneously address longitudinal data with skewness in viral response, and measurement errors and missing data in CD4 covariate. We found that it is important to take the CD4 measurement errors and missing data, and viral load with skewness into account, indicating that the JM approach may be more appropriate than the NV and TS methods.

Appendices

A. HIV dynamic models

Viral dynamic models can be formulated through a system of ordinary differential equations (ODE) (Guedj et al., 2007; Huang et al., 2006; Perelson et al., 1997; Wu and Ding, 1999). Following the notation in Huang et al., (2006) and Wu and Ding (1999), a mathematical ODE model for HIV dynamics can be written as follows by considering an infected cell compartment–productively infected cells (T_p).

$$\begin{array}{rcl} \frac{d}{dt}T_p & = & kTV_I - d_pT_p, \\ \frac{d}{dt}V_I & = & (1-\eta)d - cV_I, \\ \frac{d}{dt}V_{NI} & = & \eta d + Nd_pT_p - cV_{NI}, \end{array} \tag{A.1}$$

where V_I and V_{IN} are the concentrations of infectious virus and non-infectious virus, respectively, and T denotes the number of uninfected target cells for

HIV, which can be assumed to be a constant at the early stage of HIV treatment. To account for compartment where the protease inhibitor drugs cannot completely block the production, we consider an additional virus production term with a constant (average) rate d in the model. The parameters d_p and c are the death rates of productively infected cells and free virus, respectively. Under some reasonable assumptions and simplifications, an analytic solution for equation (A.1) can be obtained. More details on the notation and simplifications can be found in Wu and Ding (1999). Thus, two useful approximate solutions, which can be used to capture virus decay, have been proposed as follows.

$$V(t) = \exp(p_0) + \exp(p_1 - \lambda t) \tag{A.2}$$
$$V(t) = \exp(p_1 - \lambda_1 t) + \exp(p_2 - \lambda_2 t) \tag{A.3}$$

where $V(t) = V_I(t) + V_{NI}(t)$ is the total number of HIV-1 RNA copies per mL of plasma, λ, λ_1 and λ_2 are the viral decay rates before viral load rebound (Perelson et al., 1997), $\exp(p_i)$ (i=0,1,2) reflect the baseline viral load at time $t = 0$. It is generally assumed that $\lambda_1 > \lambda_2$, which assures that the model is identifiable and is appropriate for empirical studies (Wu and Ding, 1999). It is of particular interest to estimate these viral decay rates because they quantify the antiviral effect, and hence, can be used to assess the efficacy of the antiviral treatments. In estimating these decay rates, only the early segment of the viral load trajectory data before rebound can be used (Perelson et al., 1997; Wu and Ding, 1999; Wu and Wu, 2002a, 2002b).

Nonlinear mixed-effects models based on one-compartment model with one phase decay rate (A.2) and two-compartment model with two phase decay rates (A.3) are powerful tools for modeling HIV viral dynamics. Wu and Wu (2002b) have shown that they are approximately equal to capture the viral load trajectory reasonably well within some time period. Although equations (A.2) and (A.3) are widely used in HIV dynamic studies, they are only applied to the early segment of the viral load data since the viral load trajectory may change to a different shape in later stage; see Figure 1. Thus, it may not be reasonable to assume that the viral decay rate is a constant during long-term treatment such as 48 weeks in the study to be considered in this paper. In other words, equations (A.2) and (A.3) is only short-term HIV dynamic models. To model the long-term HIV dynamics, a natural extension is to assume that the viral decay rates change over time, which are either a function of time-varying covariates such as CD4 cell count or a smooth function to capture the viral load change including viral rebound. Thus, we introduce two extended function as follows.

$$V(t) = \exp(p_0) + \exp[p_1 - \lambda(t)t] \tag{A.4}$$
$$V(t) = \exp[p_1 - \lambda_1(t)t] + \exp[p_2 - \lambda_2(t)t] \tag{A.5}$$

where the decay rates $\lambda(t)$, $\lambda_1(t)$ and $\lambda_2(t)$ are either a function of time-varying CD4 cell count (covariate) and/or an unknown smooth function. Intuitively, these two models are more reasonable because they assume that the viral decay rates can vary with time as a result of drug resistance, medication adherence and other relevant clinical factors likely to affect changes in the viral load during the treatment. Therefore, all data obtained during whole study period can be used by fitting these models. We also assume that $\lambda_1(t) > \lambda_2(t)$, for all time t in order to guarantee that there is the first phase of curve decay. Equations (A.4) and (A.5) are semiparametric models because of the mechanistic structure with constant parameters and time-varying parameters to capture the time-varying effects of the treatment and over a longer period. More importantly, the semiparametric models are capable of modeling long-term viral load data of which the trajectory may vary substantially among different patients (Wu and Zhang, 2006).

B. WinBUGS code for Model I with SN distribution

```
model {
  for (i in 1:m)
  {
 # Random effects of response model
    b2[i,1]<-0
    b2[i,2]<-0
    b2[i,3]<-0
    b[i,1:3]~dmnorm(b2[i,1:3],Omega2[,])
 # Random effects of covariate model
    a3[i,1]<-0
    a3[i,2]<-0
    a3[i,3]<-0
    a[i,1:3]~dmnorm(a3[i,1:3],Omega3[,])
   }
  for(j in 1:N)
  {
 # Modelling nonignorable missing model
    logit(mu.r[j])<-psi[1]+psi[2]*y[j,1]     #y[j,1]=standardized cd4
    mu.r2[j]<-max(0.001,min(mu.r[j],0.999))
    y[j,3]~dbern(mu.r2[j])                   #y[j,3]=Indicator 0 or 1

 # Modelling true CD4 via measurement errors model
    z.star[j]<-(alpha[1]+a[y[j,2],1])+(alpha[2]+a[y[j,2],2])*y[j,4]
```

```
   +(alpha[3]+a[y[j,2],3])*y[j,4]*y[j,4]   #y[j,2]=patid; y[j,4]=time
   y[j,1]~dnorm(z.star[j],tau2)

# SNLME response model
   p1[j]<-beta[1]+b[y[j,2],1]
   p2[j]<-beta[2]+b[y[j,2],2]
   lambda1[j]<-beta[3]+beta[4]*z.star[j]
   +mu.not[1]+mu.not[2]*Z[j,2]+mu.not[3]*Z[j,3]+b[y[j,2],3]
   dm1[j]<-p1[j]
   dm2[j]<-p2[j]-lambda1[j]*y[j,4]
   dm3[j]<-exp(dm1[j])
   dm4[j]<-exp(dm2[j])
   dm5[j]<-dm3[j]+dm4[j]
   w[j]~dnorm(0,1)I(0,)
   mu[j]<-log(dm5[j])/log(10)+delta*(w[j]-0.798)  # SN distribution
 # mu[j]<-log(dm5[j])/log(10)   # Normal distribution
   y[j,5]~dnorm(mu[j], tau)     # y[j,5]=log10(rna)
   Y.pred[j]~dnorm(mu[j],tau)   # predict value

# Fitted values and Residuals
   fit[j]<-mu[j]
   Pred[j]<-Y.pred[j]
   resid[j]<-y[j,5]-fit[j]
   ssr.r[j]<-pow(resid[j],2)
   ssr.Y.obs[j]<-pow((Y.pred[j]-y[j,5]),2)
 }
   SSR<-sum(ssr.r[])         # Sum of squares of residuals
   SSR.p<-mean(ssr.Y.obs[])  # EPD

# Prior distributions of the hyperparameters
#(1) Coefficients
   for(k in 1:4){beta[k]~dnorm(0,0.01)}
   for(k in 1:3){
                alpha[k]~dnorm(0,0.01)
                mu.not[k]~dnorm(0,0.01)
                }
   for (k in 1:2){psi[k]~dnorm(0,0.01)}
#(2) Precision parameters
   tau~dgamma(0.01,0.01)
   sigma.tau<-1/tau
```

```
    tau2~dgamma(0.01,0.01)
    sigma.tau2<-1/tau2
 #(3) Variance-covariance matrix
    Omega2[1:3,1:3]~dwish(R2[,],5)
    v2[1:3,1:3]<-inverse(Omega2[,])
    Omega3[1:3,1:3]~dwish(R3[,],5)
    v3[1:3,1:3]<-inverse(Omega3[,])
 #(4) Skewness parameter
    delta~dnorm(0,0.01)
}
```

References

Arellano-Valle, R.B., Bolfarine, H., Lachos, V.H. (2005). Skew-normal linear mixed models. *Journal of Data Science* **3**, 415–438.

Arellano-Valle, R.B., Bolfarine, H., Lachos, V.H. (2007). Bayesian inference for skew-normal linear mixed models. *Journal of Applied Statistics* **34**, 663–682.

Arellano-Valle, R.B., Azzalini, A. (2006). On the unification of families of skew-normal distributions. *Scandinavian Journal of Statistics* **33**, 561-574.

Azzalini, A., and Capitanio, A. (1999). Statistical applications of the multivariate skew normal distributions. *Journal of Royal Statistical Society, Series B* **6**, 579–602.

Carroll, R.J., Ruppert, D., Stefanski, L.A., and Crainiceanu, C.M. (2006). *Measurement Error in Nonlinear Models: A Modern Perspective*, 2nd edition. London, Chapman and Hall.

Celeux, G., Forbes, F., Robert, C., and Titterington, M. (2005). Deviance information criteria for missing data models. *Bayesian Analysis* **1**(4), 651–674.

Gelman, A., Carlin, J.B., Stern, H.S., and Rubin, D.B. (2003). *Bayesian Data Analysis*. London, Chapman and Hall.

Guedj, J., Thiébaut, R., and Commenges, D. (2007). Maximum likelihood estimation in dynamical models of HIV. *Biometrics* **63**, 1198–1206.

Higgins, M., Davidian, M., and Gilitinan D.M. (1997). A two-step approach to measurement error in time-dependent covariates in nonlinear mixed-effects models, with application to IGF-I pharmacokinetics. *Journal of the American Statistical association* **92**, 436–448.

Huang, Y., Liu, D., and Wu, H. (2006). Hierarchical Bayesian methods for estimation of parameters in a longitudinal HIV dynamic system. *Biometrics* **62**, 413–423.

Huang, Y., and Dagne, G. (2010). Skew-normal Bayesian nonlinear mixed-effects models with application to AIDS studies. *Statistics in Medicine* **29**, 2384–2398.

Huang, Y., and Dagne, G. (2011). A Bayesian approach to joint mixed-effects models with a skew-normal distribution and measurement errors in covariates. *Biometrics* **67**, 260–269.

Huang, Y., and Dagne, G. (2012). Bayesian semiparametric nonlinear mixed-effects joint models for data with skewness, missing responses and measurement errors in covariates. *Biometrics* DOI: 10.1111/j.1541-0420.2011.01719.x

Hughes, J.P. (1999). Mixed effects models with censored data with applications to HIV RNA levels. *Biometrics* **55**, 625–629.

Jara, A., Quintana, F., and Martin, E.S. (2008). Linear mixed models with skew-elliptical distributions: A Bayesian approach. *Computational Statistics and Data Analysis* **52**, 5033–5045.

Lederman, M.M., Connick, E., Landay, A., Kuritzkes, D.R., Spritzler, J., Clair, S.M., Kotzin, B.L., Fox, L., Chiozzi, M.H., Leonard, J.M., Rousseau, F., Wade, M., Roe, J.D., Martinez, A., and Kessler, H. (1998). 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* **178**, 70–79.

Lunn, D.J., Thomas, A., Best, N., and Spiegelhalter, D. (2000). WinBUGS – a Bayesian modelling framework: concepts, structure, an extensibility. *Statistics and Computing* **10**, 325–337.

Perelson, A.S., Essunger, P., et al. (1997). Decay characteristics of HIV-1-infected compartments during combination therapy. *Nature* **387**, 188–191.

Sahu, S.K., Dey, D.K., and Branco, M.D. (2003). A new class of multivariate skew distributions with applications to Bayesian regression models. *The Canadian Journal of Statistics* **31**, 129–150.

Spiegelhalter, D.J., Best, N.G., Carlin, B.P., and Van der Linde, A. (2002). Bayesian measures of model complexity and fit (with Discussion). *Journal of the Royal Statistical Society, Series B* **64**, 583–639.

Wu, H., and Ding, A.A. (1999). Population HIV-1 dynamics in vivo: applicable models and inferential tools for virological data from AIDS clinical trials. *Biometrics* **55**, 410–418.

Wu, L. (2002). A joint model for nonlinear mixed-effects models with censoring and covariates measured with error. *Journal of the American Statistical*

Association **97**, 955–964.

Wu, L. (2004). Simultaneous inference for longitudinal data with detection limits and covariates measured with errors, with application to AIDS studies. *Statistics in Medicine* **23**, 1715–1731.

Wu, L., and Wu, H. (2002a). Missing time-dependent covariates in human immunodeficiency virus dynamic models. *Applied Statistics* **51**, 297–318.

Wu, H., and Wu, L. (2002b) Identification of significant host factors for HIV dynamics modeled by nonlinear mixed-effects models. *Statistics in Medicine* **21**, 753–771.

Wu H, Zhang J-T. (2006). *Nonparametric Regression Methods for Longitudinal Data Analysis*. Wiley, New Jersey.