Published in final edited form as:
Stat Methods Appt. 2014 March ; 23(1): 95–121. doi:10.1007/s10260-013-0242-7.

Jointly modeling time-to-event and longitudinal data: A Bayesian approach

Yangxin Huang,
Department of Epidemiology & Biostatistics, College of Public Health, MDC 56, University of South Florida, Tampa, FL 33612, USA, Tel.: +1-813-9748209, Fax: +1-813-9744719

X. Joan Hu, and
Department of Statistics and Actuarial Sciences, Simon Fraser University, Burnaby, BC, V5A 1S6, Canada

Getachew A. Dagne
Department of Epidemiology & Biostatistics, College of Public Health, MDC 56, University of South Florida, Tampa, FL 33612, USA, Tel.: +1-813-9748209, Fax: +1-813-9744719

Yangxin Huang: yhuang@health.usf.edu

Abstract

This article explores Bayesian joint models of event times and longitudinal measures with an attempt to overcome departures from normality of the longitudinal response, measurement errors, and shortages of confidence in specifying a parametric time-to-event model. We allow the longitudinal response to have a skew distribution in the presence of measurement errors, and assume the time-to-event variable to have a nonparametric prior distribution. Posterior distributions of the parameters are attained simultaneously for inference based on Bayesian approach. An example from a recent AIDS clinical trial illustrates the methodology by jointly modeling the viral dynamics and the time to decrease in CD4/CD8 ratio in the presence of CD4 counts with measurement errors and to compare potential models with various scenarios and different distribution specifications. The analysis outcome indicates that the time-varying CD4 covariate is closely related to the first-phase viral decay rate, but the time to CD4/CD8 decrease is not highly associated with either the two viral decay rates or the CD4 changing rate over time. These findings may provide some quantitative guidance to better understand the relationship of the virological and immunological responses to antiretroviral treatments.

Keywords

Accelerated failure time model; Dirichlet process; Semiparametric linear/nonlinear mixed-effects model; Skew-elliptical distribution; Time-to-event

1 Introduction

Many studies aim at exploring the relationship between a time-to-event out-come and a longitudinal marker. For example, the relationship of HIV viral suppression and immune restoration after a treatment has received great attention in HIV/AIDS research (e.g., Acosta et al. 2004). One is often interested in simultaneously studying the HIV dynamics and immune restoration, which may be characterized by the time to CD4/CD8 ratio decline (e.g., Pawitan and Self, 1993). Relatively little has been published about statistical analysis for the particular association of HIV viral dynamics and time trend in CD4/CD8 ratio in the presence of CD4 covariate process, with few notable exceptions such as Wu et al. (2010). The research was motivated by the AIDS study (Acosta et al. 2004) to understand within-

subject patterns of change in viral load or CD4 cell count, and to study the relationship of features of viral load and CD4 profiles with time to decrease in CD4/CD8 ratio.

Joint analysis of event times and longitudinal measures has recently attracted great attention. There have been a considerable number of statistical approaches in the literature. However, most of the published methods assume that the error terms in the models for the longitudinal response or for the measurement errors in covariates follow normal distributions due to mathematical tractability and computational convenience. This requires the variables to be "symmetrically" distributed. A violation of this assumption could lead to misleading inferences. In fact, observed data in HIV studies are often far from being "symmetric": measurement errors in covariates often arise and asymmetric patterns of observations usually occur. This motivated our development of a Bayesian joint model for non-normally distributed data with imperfectly measured covariates. Although the proposed approach applies quite generally, this article focuses on joint modeling of HIV dynamics and time to the decline of the CD4/CD8 ratio after treatment was initiated, using CD4 counts over time as a covariate process.

We employ a general Bayesian framework for joint modeling of longitudinal response and time-to-event in the presence of covariates with measurement errors, to accommodate a large class of data structures with various features. In particular, we consider an accelerated failure time (AFT) model for the time to the decline of CD4/CD8 ratio using the nonparametric Dirichlet process (DP) prior as a distribution (Ferguson 1973; Ishwaran and James 2002). The semiparametric nonlinear mixed-effects (SNLME) model with skew-elliptical (SE) distributed random errors is used for HIV dynamics. The skew-elliptical distributions include the skew-normal (SN) and skew-*t* (ST) distributions as special cases. We model the CD4 process with the linear mixed-effects (LME) model with SE distributions to account for the potential measurement errors. The modeling is motivated by the following two considerations: (i) often, one may not be confident in specifying a parametric model for event times, and (ii) normal distributions are not always appropriate for the outcomes of interest.

It is noted that the ST distribution is approximate to the SN distribution when its degrees of freedom approach infinity. Thus, we use an ST distribution to develop joint models and associated statistical methodologies, as it can be easily extended to other SE distributions such as the SN distribution. We consider multivariate ST and SN distributions introduced by Sahu et al. (2003), which are suitable for a Bayesian inference and briefly discussed in the Appendix. The rest of this paper is organized as follows. Notation and joint model setup are introduced; associated simultaneous Bayesian inference procedures for estimating all parameters of the joint models are presented in Section 2. In Section 3, we describe the data set that motivated this research, discuss the specific models for viral load response, CD4 covariate process and time to decline of CD4/CD8 ratio that are used to illustrate the proposed methodologies, and report the analysis results. We conclude the paper with a discussion in Section 4.

2 Bayesian joint models with skew distributions

Figure 1 presents the trajectories of HIV viral load, CD4 cell count and CD4/CD8 ratio of three randomly selected subjects from a recent AIDS clinical trial (Acosta et al. 2004). While the data indicate a close association of the longitudinal viral load, CD4 cell count and CD4/CD8 ratio, the data show a large variation in the association across subjects and over time within each subject. This observation together with the published researches in the area (e.g., Wu and Ding 1999; Liu and Wu 2007; Wu 2009) led us to propose the following joint models for the three components of concern.

2.1 Measurement error models with a skew-*t* distribution

Various covariate mixed-effects models were investigated in the literature (Carroll et al. 2006; Huang and Dagne 2011; Liu and Wu 2007). In this section we briefly discuss covariate mixed-effects models with random errors to have an ST distribution. For simplicity, we consider a single time-varying covariate. Let z_{ik} be the observed covariate value for individual i at time s_{ik} ($i = 1, \ldots n; k = 1, \ldots, m_i$). We allow the covariate measurement times s_{ik} to differ from the response measurement times t_{ij} ($j = 1, \ldots, n_i$) for each individual. In the presence of covariate measurement errors, we consider the following covariate LME model with an ST distribution

$$\begin{aligned} z_{ik} &= \boldsymbol{u}_{ik}^T\boldsymbol{\alpha} + \boldsymbol{v}_{ik}^T\boldsymbol{a}_i + \varepsilon_{ik} \ (\equiv z_{ik}^* + \varepsilon_{ik}), \\ \boldsymbol{\varepsilon}_i \ iid &\sim ST_{m_i,\nu_1}(-J(\nu_1)\boldsymbol{\delta}_{\varepsilon_i}, \sigma_1^2\boldsymbol{I}_{m_i}, \boldsymbol{\Delta}(\boldsymbol{\delta}_{\varepsilon_i})), \end{aligned} \quad (1)$$

where $\boldsymbol{z}_i = (z_{i1}, \ldots, z_{im_i})^T$ and z_{ik} is the covariate value for individual i at time s_{ik}, $\boldsymbol{z}_i^* = (z_{i1}^*, \ldots, z_{im_i}^*)^T$ and $z_{ik}^* = \boldsymbol{u}_{ik}^T\boldsymbol{\alpha} + \boldsymbol{v}_{ik}^T\boldsymbol{a}_i$ may be viewed as the true (but unobservable) covariate value at time s_{ik}, $\boldsymbol{u}_{ik}$ and $\boldsymbol{v}_{ik}$ are $r \times 1$ design vectors. Here $\boldsymbol{\alpha} = (a_1, \ldots, a_r)^T$ and $\boldsymbol{a}_i = (a_{i1}, \ldots, a_{ir})^T$ are unknown population (fixed-effects) and individual-specific (random-effects) parameter vectors, respectively, $\boldsymbol{\varepsilon}_i = (\varepsilon_{i1}, \ldots, \varepsilon_{im_i})^T$ follows a multivariate ST distribution with ν_1 degrees of freedom, variance σ_1^2 and $m_i \times m_i$ skewness diagonal matrix $\boldsymbol{\Delta}(\boldsymbol{\delta}_{\varepsilon_i}) = diag(\delta_{\varepsilon i1}, \ldots, \delta_{\varepsilon im_i})$ with $m_i \times 1$ skewness parameter vector $\boldsymbol{\delta}_{\varepsilon i} = (\delta_{\varepsilon i1}, \ldots, \delta_{\varepsilon im_i})^T$, and $J(\nu_1) = (\nu_1/\pi)^{1/2}[\Gamma((\nu_1 - 1)/2)/\Gamma(\nu_1/2)]$ is the function of ν_1. In particular, if $\delta_{\varepsilon i1} = \cdots = \delta_{\varepsilon im_i} \triangleq \delta_\varepsilon$, then $\boldsymbol{\delta}_{\varepsilon_i} = \delta_\varepsilon \mathbf{1}_{m_i}$ with $\mathbf{1}_{m_i} = (1, \ldots, 1)^T$ and $\boldsymbol{\Delta}(\boldsymbol{\delta}_{\varepsilon_i}) = \delta_\varepsilon \boldsymbol{I}_{m_i}$; this is the case to be considered for data analysis in Section 3. We assume that $\boldsymbol{a}_i \ iid \sim N_r(0, \boldsymbol{\Sigma}_a)$, where $\boldsymbol{\Sigma}_a$ is the unrestricted covariance matrix. Note that model (1) may be interpreted as an ST covariate measurement error model.

2.2 Skew-*t* semiparametric nonlinear mixed-effects models for response process

We now present models for response process in general forms, also appropriate in other applications. Denote the number of subjects by n and the number of measurements on the ith subject by n_i. For the response process, we consider a general SNLME model with an ST distribution and incorporate possibly mis-measured time-varying covariates. Let y_{ij} be the response value for individual i at t_{ij}, $g(\cdot)$, $\boldsymbol{d}^\dagger(\cdot)$ and $v(\cdot)$ be known parametric functions, $w(t)$ and $h_i(t)$ be unknown nonparametric smooth fixed-effects and random-effects functions, respectively, with $h_i(t)$ being iid realization of a zero-mean stochastic process. Consider

$$\begin{aligned} y_{ij} &= g(t_{ij}, \boldsymbol{\beta}_{ij}^\dagger, \varphi(t_{ij})) + e_{ij}, \ \ \boldsymbol{e}_i \ iid \sim ST_{n_i,\nu_2}(-J(\nu_2)\boldsymbol{\delta}_{e_i}, \sigma_2^2\boldsymbol{I}_{n_i}, \boldsymbol{\Delta}(\boldsymbol{\delta}_{e_i})), \\ \boldsymbol{\beta}_{ij}^\dagger &= \boldsymbol{d}^\dagger[z_{ij}^*, \boldsymbol{\beta}^\dagger, \boldsymbol{b}_i^\dagger], \ \ \varphi(t_{ij}) = v[w(t_{ij}), h_i(t_{ij})], \end{aligned} \quad (2)$$

where $\boldsymbol{\beta}_{ij}^\dagger$ is $s_1 \times 1$ individual-specific time-dependent parameter vector, $\boldsymbol{\beta}^\dagger$ is $s_2 \times 1$ population parameter vector ($s_2 \geq s_1$), $\boldsymbol{b}_i^\dagger$ ($s_3 \leq s_1$) is $s_3 \times 1$ vector of random-effects, the vector of random errors $\boldsymbol{e}_i = (e_{i1}, \ldots, e_{in_i})^T$ follows a multivariate ST distribution with degrees of freedom ν_2, variance σ_2^2 and $n_i \times n_i$ skewness diagonal matrix $\boldsymbol{\Delta}(\boldsymbol{\delta}_{e_i}) = diag(\delta_{e_i1}, \ldots, \delta_{e_in_i})$ with the $n_i \times 1$ skewness parameter vector $\boldsymbol{\delta}_{e_i} = (\delta_{e_i1}, \ldots, \delta_{e_in_i})^T$. In particular, if $\delta_{e_i1} = \cdots = \delta_{e_in_i} \triangleq \delta_e$, then $\boldsymbol{\Delta}(\boldsymbol{\delta}_{e_i}) = \delta_e \boldsymbol{I}_{n_i}$ and $\boldsymbol{\delta}_{e_i} = \delta_e \mathbf{1}_{n_i}$. In model (2), we assume that the individual-specific parameter $\boldsymbol{\beta}_{ij}^\dagger$ depends on the true (but unobserved) covariate z_{ij}^* rather than the observed covariate z_{ij}, which may be measured with substantial errors. The ST-SNLME model (2) reverts to a parametric NLME model when the nonparametric parts $w(t)$ and $h_i(t)$ are constants. To fit model (2), we may apply the regression spline method for

nonparametric functions. The working principle is briefly described as follows and more details can be found in Wu and Zhang (2002).

The main idea of 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 $\mathbf{\Psi}_p(t) = \{\psi_0(t), \psi_1(t), \ldots, \psi_{p-1}(t)\}^T$ and $\mathbf{\Psi}_q(t) = \{\psi_0(t), \psi_1(t), \ldots, \psi_{q-1}(t)\}^T$, respectively. That is,

$$\begin{aligned} w(t) &\approx w_p(t) = \sum_{l=0}^{p-1} \mu_l \psi_l(t) = \mathbf{\Psi}_p(t)^T \boldsymbol{\mu}_p, \\ h_i(t) &\approx h_{iq}(t) = \sum_{l=0}^{q-1} \xi_{il} \varphi_l(t) = \mathbf{\Phi}_q(t)^T \boldsymbol{\xi}_{iq}, \end{aligned} \quad (3)$$

where $\boldsymbol{\mu}_p$ and $\boldsymbol{\xi}_{iq}$ ($q \le p$) are the unknown vectors of fixed and random coefficients, respectively. Based on the assumption of $h_i(t)$, we regard $\boldsymbol{\xi}_{iq}$ as *iid* realizations of a zero-mean random vector. For our model, we consider natural cubic spline bases with the percentile-based knots. To select an optimal degree of regression spline and number 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).

Substituting $w(t)$ and $h_i(t)$ by their approximations $w_p(t)$ and $h_{iq}(t)$, we can approximate model (2) as follows.

$$\begin{aligned} y_{ij} &= g\left(t_{ij}, \boldsymbol{d}^{\dagger}[z_{ij}^*, \boldsymbol{\beta}^{\dagger}, \boldsymbol{b}_i^{\dagger}], v[\mathbf{\Psi}_p(t_{ij})^T \boldsymbol{\mu}_p, \mathbf{\Phi}_q(t_{ij})^T \boldsymbol{\xi}_{iq}]\right) + e_{ij} \\ &\equiv g\left(t_{ij}, \boldsymbol{d}[z_{ij}^*, \boldsymbol{\beta}, \boldsymbol{b}_i]\right) + e_{ij} \end{aligned} \quad (4)$$

where $\boldsymbol{\beta} = (\boldsymbol{\beta}^{\dagger T}, \boldsymbol{\mu}_p^T)^T$ and $\boldsymbol{b}_i = (\boldsymbol{b}_i^{\dagger T}, \boldsymbol{\xi}_{iq}^T)^T$ are fixed-effects and random-effects, respectively, and $\boldsymbol{d}(\cdot)$ is a known but possible nonlinear function. By doing so, 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 $\mathbf{\Psi}_p(t)$ and $\mathbf{\Phi}_q(t)$, we approximate the ST-SNLME model (2) by the following ST-NLME model.

$$\begin{aligned} \boldsymbol{y}_i &= \boldsymbol{g}_i(\boldsymbol{t}_i, \boldsymbol{\beta}_i) + \boldsymbol{e}_i, \;\; \boldsymbol{e}_i \;\; iid \sim ST_{n_i,\nu_2}(-J(\nu_2)\boldsymbol{\delta}_{e_i}, \sigma_2^2 \boldsymbol{I}_{n_i}, \boldsymbol{\Delta}(\boldsymbol{\delta}_{e_i})), \\ \boldsymbol{\beta}_{ij} &= \boldsymbol{d}[z_{ij}^*, \boldsymbol{\beta}, \boldsymbol{b}_i], \;\; \boldsymbol{b}_i \;\; iid \sim N_{s_4}(\mathbf{0}, \textstyle\sum_b), \end{aligned} \quad (5)$$

where $s_4 = s_3 + q$, $\boldsymbol{y}_i = (y_{i1}, \ldots, y_{in_i})^T$, $\boldsymbol{t}_i = (t_{i1}, \ldots, t_{in_i})^T$, $\boldsymbol{\beta}_i = (\boldsymbol{\beta}_{i1}, \ldots, \boldsymbol{\beta}_{in_i})^T$, $\boldsymbol{g}_i(\boldsymbol{t}_i, \boldsymbol{\beta}_i) = (g(t_{i1}, \boldsymbol{\beta}_{i1}), \ldots, g(t_{in_i}, \boldsymbol{\beta}_{in_i}))^T$ with $g(\cdot)$ being a known nonlinear function, and $\boldsymbol{\Sigma}_b$ is an unstructured covariance matrix.

2.3 Time-to-event models with an unspecified distribution

Various accelerated failure time (AFT) models for time-to-event were investigated in the literature (DeGruttola and Tu 1994; Tsiatis and Davidian 2004; Wu et al. 2010). However, the common assumption of distributions for model errors is log-normal and this assumption may lack the robustness and/or may violate the agreement with observed data. Thus, statistical inference and analysis with normal assumption may lead to misleading results. We consider AFT models with random errors to have a nonparametric prior distribution, which follow a DP prior. The event time $\mathfrak{I}_i$ is likely related to the longitudinal response and covariate processes. This association is of much interest in many practical situations. We specify the association by assuming that, conditional on the random-effects $\boldsymbol{a}_i$ and $\boldsymbol{b}_i$, the event time is related to the longitudinal processes through the random-effects that

characterize the individual-specific effects (Huang et al. 2011; Wu et al. 2010). Particularly, we consider the following AFT model.

$$\begin{aligned} \ln(\Im_i) &= \gamma_0 + \boldsymbol{\gamma}_1^T \boldsymbol{a}_i + \boldsymbol{\gamma}_2^T \boldsymbol{b}_i + \varepsilon_i, \varepsilon_i \sim G(\cdot), \\ G(\cdot) &\sim DP(\eta G_0),\ G_0 \sim N(\zeta_0, \sigma_0^2),\ \eta \sim \Gamma(\eta_{10}, \eta_{20}). \end{aligned} \quad (6)$$

where $\boldsymbol{\gamma} = (\gamma_0, \boldsymbol{\gamma}_1^T, \boldsymbol{\gamma}_2^T)^T$ are unknown parameters, $\boldsymbol{c}_i = (1, \boldsymbol{a}_i^T, \boldsymbol{b}_i^T)^T$ are the random-effects from models (1) and (5). We assume that $\boldsymbol{c}_i$ are independent of ε_i, and ε_i follows an unspecified distribution $G(\cdot)$ that is the DP prior (Ferguson 1973; Ishwaran and James 2002) with the concentration parameter η and the base measure G_0 following a Γ distribution and a specified normal distribution, respectively. Note that the parameter G_0 is the prior mean of $G(\cdot)$ and represents a guess; the other parameter of the DP prior, η, reflects the degree of closeness of $G(\cdot)$ to the prior mean G_0. Large values of η make $G(\cdot)$ very close to G_0, while small values of η allow $G(\cdot)$ to deviate from G_0. This nonparametric Bayesian analysis of specifying the distribution of ε_i is robust to misspecification of the distributional assumptions about the model error.

The AFT model (6) for time-to-event, which can be written as $\Im_i = \exp(\boldsymbol{\gamma}^T \boldsymbol{c}_i) \exp(\varepsilon_i)$, reduces to the AFT model considered in Wu et al. (2010) if $\eta = \infty$, a log-normal distribution conditional on the random-effects $\boldsymbol{c}_i$. Model (6) may be a good choice when the event times are thought to depend on individual-specific longitudinal trajectories, such as initial intercepts and slopes, or summaries of the longitudinal trajectories, and it is closely related to so-called shared parameter models (DeGruttola and Tu, 1994). As pointed out in the studies (Wu et al. 2010; Tsiatis and Davidian, 2004), in some practical situations, the event times cannot be observed but are only known as being contained in some time intervals, i.e., being interval-censored. In the AIDS study mentioned previously, for example, given that CD4 and CD8 were collected at a finite number of times, we can only know that the time to the first occurrence of CD4/CD8 decrease in a subject is between two data collection time points. The observed event time data are then $(u_i, v_i]$, $i = 1, \ldots, n$, where $(u_i, v_i]$ is the smallest observed interval containing $\Im_i$. We take u_i as the subject's latest time in the study and $v_i = \infty$ if the i^{th} subject did not experience the event of interest during the whole study period.

2.4 Simultaneous Bayesian inference for parameter estimation

In a longitudinal study, the longitudinal response and covariate processes, and the time-to-event are usually connected physically or biologically. As discussed previously, they can be jointly modeled through the shared random-effects. This joint modeling is often biologically meaningful. Statistical inference on all of the model parameters can then be made simultaneously. Thus, the underlying association can be fully addressed or captured and the uncertainty is incorporated from the estimation.

Denote the observed data by $\mathcal{D} = \{(\boldsymbol{y}_i, \boldsymbol{z}_i, u_i, v_i),\ i = 1, \ldots, n\}$, where $\boldsymbol{y}_i = (y_{i1}, \ldots, y_{in_i})^T$, $\boldsymbol{z}_i = (z_{i1}, \ldots, z_{im_i})^T$ and $(u_i, v_i]$ is the smallest observed interval containing the event time $\Im_i$. According to the properties of ST distribution, in order to specify models (1) and (5) for MCMC computation it can be shown by introducing two random variable vectors $\boldsymbol{w}_{e_i} = (w_{e_{i1}}, \ldots, w_{e_{in_i}})^T$ and $\boldsymbol{w}_{\varepsilon_i} = (w_{\varepsilon_{i1}}, \ldots, w_{\varepsilon_{im_i}})^T$ based on the stochastic representation for the ST distribution (see Appendix in detail) that $\boldsymbol{z}_i$ and $\boldsymbol{y}_i$ with respective random-effects $\boldsymbol{a}_i$ and $\boldsymbol{b}_i$ can be hierarchically formulated in conjunction with model (6) as follows.

$$\begin{gathered}\boldsymbol{z}_i|\boldsymbol{a}_i,\boldsymbol{w}_{\varepsilon_i};\boldsymbol{\alpha},\sigma_1^2,\boldsymbol{\delta}_{\varepsilon_i},\nu_1\sim t_{m_i,\nu_1+m_i}(\boldsymbol{z}_i^*+\boldsymbol{W}_{\varepsilon_i},\omega_{\varepsilon_i},\sigma_1^2\boldsymbol{I}_{m_i}),\\ \boldsymbol{y}_i|\boldsymbol{a}_i,\boldsymbol{b}_i,\boldsymbol{w}_{e_i};\boldsymbol{\alpha},\boldsymbol{\beta},\sigma_2^2,\boldsymbol{\delta}_{e_i},\nu_2\sim t_{n_i,\nu_2+n_i}(\boldsymbol{g}_i+\boldsymbol{W}_{e_i},\omega_{e_i}\sigma_2^2\boldsymbol{I}_{n_i}),\\ \Im_i|\boldsymbol{c}_i;\boldsymbol{\gamma},G_0,\eta\sim G\left(\ln(t_i)-\boldsymbol{\gamma}^T\boldsymbol{c}_i\right),\\ \boldsymbol{w}_{\varepsilon_i}|\nu_1\sim t_{m_i,\nu_1}(\boldsymbol{0},\boldsymbol{I}_{m_i})I(\boldsymbol{w}_{\varepsilon_i}>\boldsymbol{0}),\ \boldsymbol{a}_i|\textstyle\sum_a\sim N_r(\boldsymbol{0},\sum_a),\\ \boldsymbol{w}_{e_i}|\nu_2\sim t_{n_i,\nu_2}(\boldsymbol{0},\boldsymbol{I}_{n_i})I(\boldsymbol{w}_{e_i}>\boldsymbol{0}),\ \boldsymbol{b}_i|\textstyle\sum_b\sim N_{s_4}(\boldsymbol{0},\sum_b),\end{gathered}\tag{7}$$

where $\boldsymbol{W}_{\varepsilon i}=\boldsymbol{\Delta}(\boldsymbol{\delta}_{\varepsilon i})[\boldsymbol{w}_{\varepsilon i}-J(\nu_1+m_i)\boldsymbol{1}_{m_i}]$, $\boldsymbol{W}_{ei}=\boldsymbol{\Delta}(\boldsymbol{\delta}_{ei})[\boldsymbol{w}_{ei}-J(\nu_2+n_i)\boldsymbol{1}_{n_i}]$, $\omega_{\varepsilon_i}=(\nu_1+\boldsymbol{w}_{\varepsilon_i}^T\boldsymbol{w}_{\varepsilon_i})/(\nu_1+m_i)$, $\omega_{e_i}=(\nu_2+\boldsymbol{w}_{e_i}^T\boldsymbol{w}_{e_i}),(\nu_2+n_i)$, $\boldsymbol{z}_i^*=(z_{i1}^*,\ldots,z_{im_i}^*)^T$, $\boldsymbol{g}_i=\boldsymbol{g}_i(\boldsymbol{t}_i,\boldsymbol{\beta}_i)$, $t_{l,\nu}(\boldsymbol{\mu},\boldsymbol{A})$ denotes the l dimensional t-distribution with parameters $\boldsymbol{\mu}$, $\boldsymbol{A}$ and degrees of freedom ν, $I(\boldsymbol{w}>\boldsymbol{0})$ is an indicator function and $\boldsymbol{w}\sim t_{l,\nu}(\boldsymbol{0},\boldsymbol{I}_l)$ truncated in the space $\boldsymbol{w}>\boldsymbol{0}$.

Let $\boldsymbol{\theta}=\{\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\gamma},\sigma_1^2,\sigma_2^2,\sum_a,\sum_b,\nu_1,\nu_2,\boldsymbol{\delta}_{\varepsilon_i},\boldsymbol{\delta}_{e_i};i=1,\ldots,n\}$ be the collection of unknown population parameters in models (1), (5) and (6). Under Bayesian framework, we still need to specify prior distributions for all unknown parameters in the three models as follows.

$$\begin{gathered}\boldsymbol{\alpha}\sim N_r(\boldsymbol{\tau}_1,\boldsymbol{\Lambda}_1),\ \sigma_1^2\sim IG(\omega_1,\omega_2),\ \textstyle\sum_a\sim IW(\boldsymbol{\Omega}_1,\rho_1),\ \boldsymbol{\delta}_{\varepsilon_i}\sim N_{m_i}(\boldsymbol{0},\boldsymbol{\Gamma}_1),\\ \boldsymbol{\beta}\sim N_{s_5}(\boldsymbol{\tau}_2,\boldsymbol{\Lambda}_2),\ \sigma_2^2\sim IG(\omega_3,\omega_4),\ \textstyle\sum_b\sim IW(\boldsymbol{\Omega}_2,\rho_2),\ \boldsymbol{\delta}_{e_i}\sim N_{n_i}(\boldsymbol{0},\boldsymbol{\Gamma}_2),\\ \nu_1\sim Exp(\nu_{10})I(\nu_1>3),\ \nu_2\sim Exp(\nu_{20})I(\nu_2>3),\ \gamma\sim N_{s6}(\boldsymbol{\tau}_3,\boldsymbol{\Lambda}_3),\end{gathered}\tag{8}$$

where $s_5=s_2+p$, $s_6=1+s_4+r$, 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 $\boldsymbol{\Lambda}_1$, $\boldsymbol{\Lambda}_2$, $\boldsymbol{\Lambda}_3$, $\boldsymbol{\Omega}_1$, $\boldsymbol{\Omega}_2$, $\boldsymbol{\Gamma}_1$ and $\boldsymbol{\Gamma}_2$ can be assumed to be diagonal for convenient implementation.

Let $f(\cdot\,|\,\cdot)$ and $\pi(\cdot)$ be a conditional density function and a prior density function, respectively. Assume that the prior distribution of $\boldsymbol{\theta}$ to be $\pi(\boldsymbol{\theta})=\pi(\boldsymbol{\alpha})\pi(\boldsymbol{\beta})\pi(\boldsymbol{\gamma})\pi(\sigma_1^2)\pi(\sigma_2^2)\pi(\sum_a)\pi(\sum_b)\pi(\nu_1)\pi(\nu_2)\prod_i\pi(\boldsymbol{\delta}_{\varepsilon_i})\pi(\boldsymbol{\delta}_{e_i})$. 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. The joint posterior density of $\boldsymbol{\theta}$ based on the observed data can be given by

$$\begin{aligned}f(\boldsymbol{\theta}|\mathscr{D})\propto\{\prod_i^n\int\int f(\boldsymbol{y}_i|\boldsymbol{a}_i,\boldsymbol{b}_i,\boldsymbol{w}_{e_i};\boldsymbol{\alpha},\boldsymbol{\beta},\sigma_2^2,\boldsymbol{\delta}_{e_i},\nu_2)f(\boldsymbol{w}_{e_i}|\boldsymbol{w}_{e_i}>\boldsymbol{0})\\ \times f(\boldsymbol{z}_i|\boldsymbol{a}_i,\boldsymbol{w}_{\varepsilon_i};\boldsymbol{\alpha},\sigma_1^2,\boldsymbol{\delta}_{\varepsilon_i},\nu_1)f(\boldsymbol{w}_{\varepsilon_i}|\boldsymbol{w}_{\varepsilon_i}>\boldsymbol{0})\\ \times F^*(u_i,v_i|\boldsymbol{c}_i;\boldsymbol{\gamma},G_0,\eta)f(\boldsymbol{a}_i|\textstyle\sum_a)f(\boldsymbol{b}_i|\sum_b)d\boldsymbol{a}_i d\boldsymbol{b}_i\}\pi(\boldsymbol{\theta}),\end{aligned}\tag{9}$$

where $F^*(u_i,v_i|\boldsymbol{c}_i;\boldsymbol{\gamma},G_0,\eta)=G\left(\ln(v_i)-\boldsymbol{\gamma}^T\boldsymbol{c}_i\right)-G\left(\ln(u_i)-\boldsymbol{\gamma}^T\boldsymbol{c}_i\right)$. In general, the integrals in (9) 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, MCMC procedures can be used to sample based on (9) using the Gibbs sampler along with the Metropolis-Hastings (M-H) algorithm. An important advantage of the above representations based on the hierarchical models is that they can be very easily implemented using the freely available WinBUGS software (Lunn et al., 2000) and the computational effort is equivalent to the one necessary to fit the normal version of the model. Note that when using the WinBUGS software, it is not necessary to specify the full conditional distributions explicitly or proportional functions of the density functions of full conditional distributions for parameters. Although their derivations are straightforward by working the complete joint

posterior, some cumbersome algebra will be involved. Thus, we omit those here to save space.

3 An AIDS clinical study

The aforementioned HIV/AIDS clinical trial (Acosta et al. 2004) consisted of 44 HIV-infected patients, who were treated with a potent antiretroviral regimen. HIV RNA viral load was measured in copies/mL at study days 0, 7, 14, 28, 56, 84, 112, 140 and 168 of follow-up. Covariates such as CD4 and CD8 cell counts were also measured throughout the study on a similar scheme. The nucleic acid sequence-based amplification assay was used to measure RNA viral load, with a lower limit of quantification of 50 copies/mL. The RNA viral load measures below this limit are not considered reliable. Therefore, we replaced them by 25 copies/mL as in Acosta et al. (2004). The exact day of viral load measurement (not predefined study day) was used to compute study day in our analysis. We present below an application of the proposed approach in Section 2. The natural logarithmic transformation of HIV RNA copies is used in the analysis.

3.1 Model specifications

Viral dynamic models can be formulated through a system of ordinary differential equations (Huang et al. 2006; Wu and Ding 1999). Based on biological and clinical arguments, a biphasic model with both the first- and second-phase viral decay rates as constants can be approximately offered to describe HIV viral dynamics (Wu and Ding 1999). The first- and second-phase viral decay rates may represent the minimum turnover rate of productively infected cells and that of latently or long-lived infected cells, respectively. It is of particular interest to estimate these two viral decay rates because they quantify the antiviral effect and, hence, can be used to assess the efficacy of the antiviral treatment. In estimating these decay rates, only the early segment of the viral load trajectory data have been used (Wu and Ding 1999). Since the viral load trajectory may change to different shapes in the late stages, it may not be reasonable to assume that the second-phase decay rate remains constant during long-term treatment.

To model the long-term viral load responses, a semiparametric biexponential model can be constructed as follows (Wu and Zhang 2002).

$$y(t)=\ln\{V(t)\}+e(t)=\ln\{e^{p_1-\lambda_1 t}+e^{p_2-\lambda_2(t)t}\}+e(t), \quad (10)$$

where $V(t)$ is the plasma HIV-1 RNA levels (viral load) at time t; λ_1 and $\lambda_2(t)$ are called the first- and second-phase viral decay rates, respectively. The second-phase decay rate $\lambda_2(t)$ is an unknown smooth function. Model (10) reduces to that considered by Wu and Ding (1999) and Huang and Dagne (2011) when $\lambda_2(t)$ is constant. Intuitively, model (10) is more reasonable because it assumes that the decay rate can vary with time as a result of drug resistance, pharmacokinetics, drug adherence and other relevant clinical factors. Therefore, all data observed during antiretroviral treatment can be used to fit model (10). This is a semiparametric model because of the mechanistic structure (two-exponential) with constant parameters (λ_1, p_1, p_2) and a time-varying parameter ($\lambda_2(t)$) to capture the various viral load trajectories including viral load rebound over a long period. This semiparametric model preserves compartmental mechanistic interpretation of the original parametric model under the biexponential form. In particular, the turnover rate of productively infected cells, λ_1, can still be estimated. In fact, by including long-term viral load data, the estimate of λ_1 can be more accurate and reasonable compared with those obtained in previous studies such as Wu and Ding (1999) and among others, after excluding long-term viral load data for modeling and analysis by some *ad hoc* rules (that is, the screening and inclusion of viral load data are

quite arbitrary). The estimate of $\lambda_2(t)$ provides not only an approximate turnover rate over time of long-lived/latently infected cells at the early stage of treatment as in the standard parametric model, but more importantly, it describes how it may change over a long treatment period as driven by, presumably, drug resistance, non-compliance and other clinical determinants. Most importantly, the semiparametric model is capable of modeling long-term viral load data of which the trajectory may vary substantially among different patients.

Models for the covariate process are needed in order to incorporate measurement errors in covariates. CD4 measures are known with nonnegligible errors and ignoring covariate measurement errors can lead to severely misleading results in a statistical inference. In this AIDS study, CD4 measurements may be missed at viral load measurement times due mainly to a different CD4 measurement scheme as designed in the study (for example, CD4 measurements were missed at day 7 displayed in Figure 1). Thus, we assume that the missing data in CD4 are missing at random (MAR) in the sense of Rubin (Rubin 1976), such that the missing data mechanism may be ignored in the analysis. With CD4 measures collected over time from the AIDS study, we may model the CD4 process to partially address the measurement errors (Liu and Wu 2007; Huang and Dagne 2011). However, the CD4 trajectories are often complicated. To model the covariate CD4 process, given 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 $\boldsymbol{u}_{ij}=\boldsymbol{v}_{ij}=(1, t_{ij}, \ldots, t_{ij}^{r-1})^T$ and focus on linear (r = 2), quadratic (r = 3) and cubic (r = 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 ST-LME model for the observed CD4 covariate process.

$$z_{ij}=(\alpha_1+a_{i1})+(\alpha_2+a_{i2})t_{ij}+(\alpha_3+a_{i3})t_{ij}^2+\varepsilon_{ij}, \quad (11)$$

where $\boldsymbol{\varepsilon}_i=(\varepsilon_{i1}, \ldots, \varepsilon_{in_i})^T \sim ST_{n_i,\nu_1}(-J(\nu_1)\delta_\varepsilon \mathbf{1}_{n_i}, \sigma_1^2 \boldsymbol{I}_{n_i}, \delta_\varepsilon \boldsymbol{I}_{n_i})$, $z_{ij}^*=(\alpha_1+a_{i1})+(\alpha_2+a_{i2})t_{ij}+(\alpha_3+a_{i3})t_{ij}^2$, $\boldsymbol{\alpha}=(\alpha_1, \alpha_2, \alpha_3)^T$ is population (fixed-effects) parameter vector and individual-specific random-effects $\boldsymbol{a}_i=(a_{i1}, a_{i2}, a_{i3})^T \sim N_3(\mathbf{0}, \boldsymbol{\Sigma}_a)$. In addition, in order to avoid too small or large estimates which 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) such that the time scale is between 0 and 1.

Because viral load is measured on each subject repeatedly over the study period, the measurements obtained from the same subject may be correlated, but they are assumed independent between patients. One powerful tool available to handle such longitudinal data is NLME modeling, in which within-subjects and between-subjects variations are both considered. In the present context of semiparametric model, it is natural to introduce the ST-SNLME model in conjunction with the HIV dynamic model (10) as follows.

$$\begin{gathered} y_{ij}=\ln\{e^{p_{i1}-\lambda_{ij1}t_{ij}}+e^{p_{i2}-\lambda_{ij2}(t_{ij})t_{ij}}\}+e_{ij}, \\ \beta_{ij1}\equiv p_{i1}=\beta_1+b_{i1}, \quad \beta_{ij2}\equiv \lambda_{ij1}=\beta_2+\beta_3 z_{ij}^*+b_{i2}, \\ \beta_{ij3}\equiv p_{i2}=\beta_4+b_{i3}, \quad \beta_{ij4}(t_{ij})\equiv \lambda_{ij2}(t_{ij})=w(t_{ij})+h_i(t_{ij}), \end{gathered} \quad (12)$$

where y_{ij} is the natural logarithmic transformation of the viral load for the i^{th} subject at time t_{ij} (i = 1, 2, …, n, j = 1, 2, …, n_i) and

$\boldsymbol{e}_i=(e_{i1},\ldots,e_{in_i})^T\sim ST_{n_i,\nu_2}(-J(\nu_2)\delta_e\mathbf{1}_{n_i},\sigma_2^2\boldsymbol{I}_{n_i},\delta_e\boldsymbol{I}_{n_i}),\boldsymbol{\beta}_{ij}=(\beta_{ij1},\beta_{ij2},\beta_{ij3},\beta_{ij4})^T$, $\boldsymbol{\beta}_{ij}$ = $(\beta_{ij1}, \beta_{ij2}, \beta_{ij3}, \beta_{ij4})^T$ is individual parameters for the i^{th} subject at time t_{ij}. Note that the ln-transformation of the viral load is taken here to stabilize the variation of measurement errors and to speed up estimation algorithm. Model (12) is a natural extension of model (10) that accommodates time-varying covariate CD4 into the first-phase decay rate and specifies unknown nonparametric smooth function for the second-phase decay rate for capturing long-term viral load trajectories with different shapes including viral load rebound.

We employ the linear combinations of natural cubic splines with percentile-based knots to approximate the nonparametric functions $w(t)$ and $h_i(t)$. Following studies in (Liu and Wu 2007; Wu and Zhang 2002), we set $\psi_0(t) = \phi_0(t) \equiv 1$ and take the same natural cubic splines in the approximations (3) with $q \le p$. The values of p and q are determined based on the AIC/BIC which suggest the following function for $\beta_{ij4}(t)$ with $p = 3$ and $q = 1$ in the model (12) below.

$$\beta_{ij4}(t_{ij}) \approx \beta_5+\beta_6\psi_1(t_{ij})+\beta_7\psi_2(t_{ij})+b_{i4}. \quad (13)$$

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

Let $\Im_i$ be time to the first decline in the CD4/CD8 ratio for the i^{th} subject. We are interested in the association of time to immune suppression of the individual-specific initial viral decay rates and the true CD4 trajectory, which are characterized by the random-effects in the viral load response and CD4 covariate models. We may view the (unobservable) random-effects as error-free covariates in time-to-event models. The associated inference can be computationally challenging when joint models consist of (nonlinear) longitudinal models and the semiparametric Cox proportional hazards model for time-to-event, a commonly used failure time model (Tsiatis and Davidian 2004), especially when the event times are interval censored. To focus on the primary issues discussed in the article, we consider the following AFT model for time to first decline of the CD4/CD8 ratio.

$$\ln(\Im_i)=\boldsymbol{\gamma}^T\boldsymbol{c}_i+\varepsilon_i,\ \ \varepsilon_i\sim G(\cdot), \quad (14)$$

where $\boldsymbol{c}_i = (1, a_{i1}, a_{i2}, b_{i2}, b_{i4})^T$ with unknown coefficients $\boldsymbol{\gamma} = (\gamma_0, \gamma_1, \gamma_2, \gamma_3, \gamma_4)^T$ and we assume ε_i follows an unspecified distribution $G(\cdot)$. In model (14), the random-effects b_{i2} and b_{i4} represent individual variations in the first- and second-phase viral decay rates, respectively, so they may be predictive for event times. While b_{i1} and b_{i3} represent variations in the baseline viral loads, they do not appear to be highly predictive of event times, so they are excluded from the model to reduce the number of parameters. The random-effects a_{i1} and a_{i2} capture the main features of individual CD4 trajectories (such as initial intercept and slope) and thus, are included in the model. Model (14) is a parametric AFT model with random-effects as covariates and appears to fit the current dataset reasonably well. It offers the advantages of easy interpretation and robustness against neglected covariates (Wu et al. 2010; Hougaard 1999). Model (14) offers the following advantages: (i) the random-effects in the covariate model summarize the history of the covariate process, with individual-specific intercepts and rates of change, and the summary quantities are likely better predictors than the covariates at several particular times; (ii) the random-effects in the response model summarize individual variations in the first- and second-phase viral decay rates to predict time to immune suppression; (iii) the link between the three models is made clear by the shared random-effects; (iv) the model error is assumed to be more flexible and unspecified distribution that has DP prior; (v) it is easy to

implement. Given the CD4 and CD8 evaluation scheme, the times to the CD4/CD8 decrease in the analysis were interval-censored. The median of the observed interval length was 15 days, while three of forty-four subjects in the study did not experience a CD4/CD8 decrease.

3.2 Data analysis

In this section, we analyze the AIDS data set described in Section 3.1 to illustrate our methodologies in conjunction with the three models (11), (12) and (14) for the specified prior distributions (8). As discussed previously, the viral load (in ln scale) and (standardized) CD4 cell count in this data set clearly indicate their asymmetric nature. Thus, it appears adequate fitting models with a skew distribution to the data set. With this information, the following three statistical models with different specification of distributions for the model errors based on both the SNLME response model (12) for viral load response and the LME covariate model (11) for CD4 process in conjunction with the AFT model (14) with an unspecified distribution for the time to CD4/CD8 decline are employed to compare their performance.

- **Model I:** A model with independent multivariate ST distributions of random errors for both the response model (12) and the covariate model (11).
- **Model II:** A model with independent multivariate SN distributions of random errors for both the response model (12) and the covariate model (11).
- **Model III:** A model with independent multivariate normal distributions of random errors for both the response model (12) and the covariate model (11).

We will investigate the following *three scenarios*. First, since a normal distribution is a special case of the SN distribution when skewness parameter is zero, while the ST distribution reduces to the SN distribution when the degrees of freedom approach infinity, we will investigate 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 estimate the model parameters by using the 'naive' method, which does not separate the measurement errors from the true CD4 values. On other words, the 'naive' method (denoted by NM) only uses the observed CD4 values z_{ij} rather than true (unobserved) CD4 values z_{ij}^* in the response model (12), and we use it as a comparison to the joint modeling (JM) method proposed in Section 2. This comparison attempts to investigate how the measurement errors in CD4 influence modeling results. Third, we investigate the JM approach based on 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 model (12), to evaluate how the semiparametric NLME model (12) contributes to modeling results in comparison with the parametric NLME model.

In order to carry out the Bayesian inference, we must specify the values of the hyper-parameters in the prior distributions. In the Bayesian analysis, we only need to specify the priors at the population level. We take weakly informative prior distribution for the parameters in the models. 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 $\boldsymbol{\gamma}$. (ii) For the variances σ_1^2 and σ_2^2 we assume a limiting 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 $\boldsymbol{\Sigma}_b$ are taken to be inverse Wishart distributions $IW(\boldsymbol{\Omega}_1, \rho_1)$ and $IW(\boldsymbol{\Omega}_2, \rho_2)$ with covariance matrices $\boldsymbol{\Omega}_1 = diag(0.01, 0.01, 0.01)$, $\boldsymbol{\Omega}_2 = diag(0.01, 0.01, 0.01, 0.01)$ and $\rho_1 = \rho_2 = 4$, respectively. (iv) The degrees of freedom parameters ν_1 and ν_2 follow truncated exponential distribution with $\nu_{10} = \nu_{20} = 0.5$. (v) For each of the skewness parameters δ_e and δ_ε, we choose independent

normal distribution $N(0, 100)$, where we assume that $\boldsymbol{\delta}_{e_i} = \delta_e \mathbf{1}_{n_i}$ and $\boldsymbol{\delta}_{\varepsilon_i} = \delta_\varepsilon \mathbf{1}_{m_i}$ to indicate that we are interested in skewness of both overall viral load data and overall CD4 cell count data. (vi) The hyper-parameters for the DP prior are assigned by $\eta_{10} = 0.5$, $\eta_{20} = 0.01$, $\zeta_0 = 0$ and $\sigma_0^2 = 100$.

The MCMC sampler was implemented using WinBUGS software (Lunn et al. 2000), and the program codes are available from the authors upon request. In particular, the MCMC scheme for drawing samples from the posterior distributions of all parameters in the three joint models is obtained by iterating between the following two steps: (i) Gibbs sampler is used to update $\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \sigma_1^2, \sigma_2^2, \boldsymbol{\Sigma}_a, \boldsymbol{\Sigma}_b, \nu_1, \nu_2, \delta_\varepsilon$ and δ_e; (ii) we update $\mathbf{b}_i$ and $\mathbf{a}_i$ ($i = 1, 2, \cdots, n$) using the M-H algorithm. After collecting the final MCMC samples, we are able to draw statistical inference for the unknown parameters. Specifically, we are interested in the posterior means and quantiles. See the articles (Lunn et al. 2000; Huang et al. 2006) 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 implementation is applied to the actual clinical data, convergence of the generated samples is assessed using standard tools within WinBUGS software such as trace plots and Gelman-Rubin (GR) diagnostics (Gelman and Rubin, 1992; Ntzoufras, 2009). Figure 2 shows the dynamic version of GR diagnostics as obtained from the WinBUGS software for the representative parameters where the three curves are given: the middle and bottom curves below the dashed horizontal line (indicated by the value one) represent the pooled posterior variance ($\hat{V}$, green color) and average within-sample variance (W, blue color), respectively, and the top curve represents their ratio ($\hat{R}$, red color), which is associated with the potential scale reduction factor (Gelman and Rubin, 1992; Brooks and Gelman, 1998). Note that in WinBUGS, these measures of posterior variability are estimated based on the widths of the 80% posterior credible intervals (see Ntzoufras, 2009 in detail). It is seen that $\hat{R}$ tends to 1, and $\hat{V}$ and W will stabilize as the number of iterations increase indicating that the algorithm has approached convergence. With the GR convergence diagnostics observed, we propose that every 20th MCMC sample is retained from the next 200,000 after an initial number of 50,000 burn-in iterations of a single chain of length 250,000. Thus, we obtain 10,000 samples of targeted posterior distributions of the unknown parameters for statistical inference. Along with this sampling procedure, a sensitivity analysis has been performed to ensure that the data, rather than priors, dominate the posterior estimates (see Section 4 for details). We also check the k-lag serial correlation of the samples for each parameter to diagnose independence of MCMC samples. We graphically checked the last 500 samples drawn from MCMC sampling scheme for each parameter (plot not shown here) and found that the consecutive samples move randomly towards different directions, indicating that MCMC is not “sticky” and MCMC samples are independent for each parameter, suggesting the convergence to the stationary distribution.

3.3 Analysis results

The Bayesian SNLME joint modeling approach in conjunction with the three models for the viral load response, the CD4 covariate process and the time to first decline of CD4/CD8 ratio with different specifications of model distribution 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 was not completely satisfactory due to unusual data fluctuation patterns for these patients, particularly in 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 and normal Q-Q plots (bottom panel) from Models I, II and III are presented in Figure 3. It was

seen from Figure 3 (top panel) that the model where the random error is assumed to have ST and SN distributions gave a similar performance, but they provided a better fit to the observed data, compared with the model where the random error is assumed to have normal distribution. This result can also be explained by examining the Q-Q plots of the residuals (bottom panel) that all plots show the existence of outliers, but it is clearly seen that Models I and II only have few negative outliers and thus, provide a better goodness-of-fit to the data than Model III. Further, Model I betetr accounts for heaviness in the tail as well as skewness. This finding is further confirmed by their standardized residual sum of squares (RSS), which are 219.8 (ST random error), 309.3 (SN random error) and 410.4 (normal random error).

The population posterior mean (PM), corresponding standard deviation (SD) and 95% credible interval for fixed-effects parameters based on Models I, II and III with two methods (joint modeling 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 (12), the findings, particularly for the fixed-effects (β_2, β_3) which are coefficient parameters of the first-phase viral decay rate λ_1, show that these estimates are different from zero for all three models since the 95% credible intervals do not contain zero. Nevertheless, for the coefficient of CD4 covariate β_3, its estimate is significantly positive. That means CD4 has a 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 during treatment. Further, the estimate of β_3 based on Model I is larger than those based on Models II and III. There are marked differences in posterior means of the variance parameter σ_2^2 across these Models considered, which are 0.01, 0.07 and 1.25 for Models I, II and III, respectively. The estimated values of the variance in both Model I and Model 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. Furthermore, Model I has the least estimate of σ_2^2 since it considers heaviness in tails as well as skewness. That is, the standard error of Model I is smaller than that of Model II, indicating that Model I with longer tails seems to produce more accurate posterior estimate than Model II. Thus, Model I, which is based on ST distribution for the model error, is the most preferred model to use in modeling viral load responses. The estimates of the skewness parameter (δ_e) of Models I and II are 1.17 with 95% credible intervals (1.00, 1.43) and 1.87 with 95% credible intervals (1.58, 2.12), respectively. This finding suggests that there is a positive and significant skewness in the data and confirms the fact that the distribution of the original data is skewed even after taking the ln-transformation. Thus, incorporating skewness parameter in modeling the data is recommended. Furthermore, the estimate (1.17) of the skewness parameter based on Model I with ST distribution is less than that (1.87) of Model II with SN distribution. This may be due to the fact that an additional parameter, the degrees of freedom ν_2, for heaviness in the tails was estimated to have 3.28 trading-off the effect of skewness.

For parameter estimates of the CD4 covariate model (4), the estimate of the coefficient α_2 in Model I is slightly smaller than that in Model II. Its estimate is significantly different from zero with a positive value. This finding suggests that there is a positively linear relationship between CD4 cell count and measurement time, and this result may be explained by the fact that CD4 cell count may increase during treatment, and in turn, indicates an overall improvement of AIDS progression in this AIDS trial study. There are differences in the posterior means of the variance σ_1^2, which are 0.05, 0.07, and 0.13 for Model I, II and III, respectively. The estimated values in both Models I and II are much smaller than that in Model III because the former models take into account skewness of the data while the latter

does not. Furthermore, Model I has the least estimate of σ_1^2 since it considers heaviness in tails as well as skewness.

The estimates for the parameters in the AFT model (14) do not directly show that time to CD4/CD8 decrease is highly associated with either the two viral decay rates or the CD4 changing rates over time, which is different from what was anticipated. This finding is consistent with the study by Wu et al. (2010).

For comparison, we employed the NM based on Model I to estimate the model parameters presented in Tables 1 and 2 using the observed CD4 and ignoring the CD4 measurement error in model (12). The difference between the naive estimates and the joint modeling estimates, depending on whether or not to ignore potential CD4 measurement error in conjunction with model (12), indicates that CD4 measurement error can not be ignored in the analysis.

We also considered the joint modeling approach for a parametric NLME model, denoted by PJM, in comparison with the semiparametric NLME model developed in this paper. We find that the parametric model may lead to possibly underestimates and/or overestimates (see, e.g., estimates of β_2, β_3, β_5 and β_6). Thus, the semiparametric model offers possibly more reliable estimates of some parameters, which may lead to better understanding of the HIV viral dynamics and the treatment efficacy.

4 Concluding discussion

We present a Bayesian joint modeling for three components (response, covariate and time-to-event processes) linked through the random-effects that characterize the underlying individual-specific longitudinal processes. We consider SNLME models with an asymmetric (ST or SN) distribution in comparison with a symmetric (normal) distribution. The approach is applied to jointly model HIV dynamics and time to decrease in CD4/CD8 ratio in the presence of the CD4 covariate process with measurement error to provide a tool to assess antiretroviral treatment and to monitor disease progression. Among the models considered in the application, Model I with ST distribution for model error was found to be favorable.

The estimated results presented in Table 1 based on (best) Model I indicate that the population CD4 trajectory may be approximated by the quadratic polynomial $\hat{z}(t) = 228.57(-0.10+0.52t - 0.14t^2)+375.46$, where $\hat{z}(t)$ is in the original CD4 scale. Figure 4 shows the estimated first- and second-phase rates ($\hat{\lambda_1}$ and $\hat{\lambda_2}(t)$) of change in viral load based on Model I. Thus, the population viral load process may be approximated by $\hat{V}(t) = \exp(6.09 - \hat{\lambda_1}t)+\exp(2.13-\hat{\lambda_2}(t)t)$. Since the first-phase viral decay rate (λ_1) is closely associated with the true CD4 values (due to significant estimate of β_3), the viral load change $V(t)$ may be highly associated with the true CD4 values. The simple approximation considered here may provide new scientific insights into the true association.

The analysis results indicate that the first-phase and second-phase time varying viral decay rates are always positive and negative in counterpart. The results in Figure 4 indicate that the first-phase (the second-phase) decay rate increases (decreases) at the early stage before approximately day 80 and then decreases (increases) at the late stage. This finding is biologically meaningful and may reflect viral load trajectory on the rapid decay phase and then a slow growth phase on population level. The true CD4 process has a significantly positive effect on the first-phase viral decay rate; this finding confirms that the CD4 covariate may be a significant predictor of the first-phase viral decay rate during the treatment process. More rapid increases in CD4 cell count may be associated with faster viral decay in the early stages. This may be explained by the fact that higher CD4 cell counts

suggest a higher turnover rate of lymphocyte cells, which may cause a positive correlation between viral decay and CD4 cell count. In addition, the estimate of within-subject variance (σ_2^2) across the three models considered is 0.01 for Model I, 0.07 for Model II and 1.25 for Model III; it indicates that a gain in significant efficiency for the skew model relative to the normal model is observed for its estimation. This is expected because high variability, heaviness of the tails and skewness are interrelated to a certain extent.

For viral load response, the skewness parameter δ_e and the degrees of freedom ν_2 based on Model I are estimated as 1.17 and 3.28. This confirms the positive skewness with heavy tail of the viral load. It may suggest that accounting for significant skewness and heaviness in the tails be required to model the data when data exhibit skewness with heavy tails. For the covariate model, the estimate of skewness parameter δ_ε and the degrees of freedom ν_1 based on Model I are 0.13 and 3.25 indicating slightly positive skewness but not significant since the 95% credible interval contains zero. This may be explained by the fact that an additional parameter ν_1 for heaviness in the tails was estimated to have 3.25 lessening the effect of skewness. In addition, the estimated results of the parameters in the time-to-event model (14) based on (best) Model I suggest that the time to first decline of CD4/CD8 ratio is not highly associated with either the two viral decay rates or the CD4 changing rates over time. This finding is consistent with that in the study by Wu et al. (2010) in which further explanations are provided. It is worth noting, given that the complex structure of the models is considered, there is the potential for overfitting due to the relatively small number of subjects involved in this study.

In order to examine the sensitivity of parameter estimates to the prior distributions, we conducted a limited sensitivity analysis using the uniform prior distribution $U(0, 100)$ or the half-Cauchy prior distribution with scale parameter of 25 for standard deviation scales σ_1 and σ_2 (Gelman, 2006) instead of inverse gamma prior distribution (used in this paper) for variance parameters σ_1^2 and σ_2^2; and we refitted data based on Model I. The results are summarized in Tables 1 and 2. It was shown from the results of sensitivity analysis that the conclusions with different prior distributions are in agreement. Thus, the findings from our analysis remain unchanged.

Modeling skewness by modifying well-known distributions is a topic that has received much attention over the past years. In the presence of skewness in the longitudinal data and measurement errors in covariates, we propose a robust Bayesian approach to an SNLME joint model based on the ST distribution as a powerful tool to handle time-to-event data and longitudinal data with asymmetric and discrepant behaviors in repeated measurements. The proposed methods enhance the modeling flexibility and allow practitioners to analyze longitudinal and event time data in a wide variety of considerations. In addition, the proposed joint modeling approach can be easily implemented using the publicly available WinBUGS package. This makes our approach quite powerful and accessible to practicing statisticians in the field. It is noted that further analysis to simply use the unreliable observations (actually observed values) below limit of quantification or to adopt more advanced imputation methods, instead of such values imputed by half of limit of quantification, may be conducted for inference, but results of parameter estimates may be interpreted differently. In addition, there are certain limitations to note in our study. For example, the current study is not intended to consider nonignorable missing data and data with detection limit problems. These complicated problems are beyond the focus of this article, but a further study may be warranted. A final issue to note is that in our application, both time-to-event and NLME response models are related to CD4 cell count, but CD4 cell count was incorporated into these two models for different purposes. NLME response models accommodate time-varying CD4 cell count as a covariate into the first-phase decay rate to predict viral trajectories in the early period, while the CD4 cell count was used to

determine time to first decline in the CD4/CD8 ratio which served as an outcome in the time-to-event model. Since time to decrease in the CD4/CD8 ratio is closely associated with HIV viral rebound, it is thus related to a long-term antiviral response in the later period.

Acknowledgments

The authors are extremely grateful to the Editor, an Associate Editor and an anonymous reviewer for their insightful comments and suggestions that led to a marked improvement of the article. This research was partially supported by NIAID/NIH grant R03AI080338 and USF Proposal Enhancement grant to Huang, Natural Sciences and Engineering Research Council of Canada discovery grant to Hu, and NIMH/NIH grant R01MH040859-22 to Dagne.

References

1. Acosta EP, Wu H, Walawander A, Eron J, Pettinelli C, Yu S, Neath D, Ferguson E, Saah AJ, Kuritzkes DR, Gerber JG. for the Adult ACTG 5055 Protocol Team . Comparison of two indinavir/ritonavir regimens in treatment-experienced HIV-infected individuals. Journal of Acquired Immune Deficiency Syndromes. 2004; 37:1358–1366. [PubMed: 15483465]
2. Arellano-Valle RB, Genton MG. On fundamental skew distributions. Journal of Multivariate Analysis. 2005; 96:93–116.
3. Arellano-Valle RB, Bolfarine H, Lachos VH. Bayesian inference for skew-normal linear mixed models. Journal of Applied Statistics. 2007; 34:663–682.
4. Azzalini A, Capitanio A. Distributions generated by perturbation of symmetry with emphasis on a multivariate skew *t* distributions. Journal of Royal Statistical Society, Series B. 2003; 65:367–389.
5. Brooks SP, Gelman A. General methods for monitoring convergence of iterative simulations. Journal of Computational and Graphical Statistics. 1998; 7(4):434–455.
6. Carroll, RJ.; Ruppert, D.; Stefanski, LA.; Crainiceanu, CM. Measurement Error in Nonlinear Models: A Modern Perspective. 2. Chapman and Hall; London: 2006.
7. DeGruttola V, Tu XM. Modeling progression of CD4-lymphocyte count and its relationship to survival time. Biometrics. 1994; 50:1003–1014. [PubMed: 7786983]
8. Ferguson TS. A Bayesian analysis of some nonparametric problems. Annals of Statistics. 1973; 1:209–230.
9. Gelman A, Rubin DB. Inference from iterative simulation using multiple sequences. Statistical Science. 1992; 7(4):457–511.
10. Gelman A. Prior distributions for variance parameters in hierarchical models. Bayesian Analysis. 2006; 1:515–533.
11. Hougaard P. Fundamentals of survival data. Biometrics. 1999; 55:13–22. [PubMed: 11318147]
12. 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]
13. 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]
14. Huang Y, Dagne G, Wu L. Bayesian inference on joint models of HIV dynamics for time-to-event and longitudinal data with skewness and covariate measurement errors. Statistics in Medicine. 2011; 30:2930–2946. [PubMed: 21805486]
15. Ishwaran H, James L. Dirichlet process computing in finite normal mixtures: smoothing and prior information. Journal of Computational and Graphical Statistics. 2002; 11:508–532.
16. Liu W, Wu L. Simultaneous inference for semiparametric nonlinear mixed-effects models with covariate measurement errors and missing responses. Biometrics. 2007; 63:342–350. [PubMed: 17688487]
17. Lunn DJ, Thomas A, Best N, Spiegelhalter D. WinBUGS – a Bayesian modelling framework: concepts, structure, an extensibility. Statistics and Computing. 2000; 10:325–337.
18. Ntzoufras, I. Bayesian Modeling Using WinBUGS. Wiley; New Jersey: 2009.

19. Pawitan Y, Self S. Modeling disease marker processes in AIDS. Journal of the American Statistical Association. 1993; 88:719–726.
20. Rubin DB. Inference and missing data. Biometrika. 1976; 63:581–592.
21. 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.
22. Tsiatis AA, Davidian M. Joint modeling of longitudinal and time-to-event data: an overview. Statistica Sinica. 2004; 14:809–834.
23. Wu H, Zhang J-T. The study of long-term HIV dynamics using semi-parametric non-linear mixed-effects models. Statistics in Medicine. 2002; 21:3655–3675. [PubMed: 12436462]
24. Wu H, Ding AA. Population HIV-1 dynamics in vivo: applicable models and inferential tools for virological data from AIDS clinical trials. Biometrics. 1999; 55:410–418. [PubMed: 11318194]
25. Wu, L. Mixed Effects Model for Complex Data. Chapman and Hall; London: 2009.
26. Wu L, Liu W, Hu XJ. Joint inference on HIV viral dynamics and immune suppression in presence of measurement errors. Biometrics. 2010; 66:327–335. [PubMed: 19673859]

Appendix. Multivariate skew distributions

Different versions of the multivariate skew–normal(SN) and skew-*t* (ST) distributions have been considered and used in the literature (Arellano-Valle and Genton 2005; Arellano-Valle et al. 2007; Azzalini and Capitanio 2003; Sahu et al. 2003 and others). A new class of distributions by introducing skewness in multivariate elliptically distributions were developed in publication (Sahu et al. 2003). The class, which is obtained by using transformation and conditioning, contains many standard families including the multivariate SN and ST distributions. For completeness, this appendix briefly summarizes the multivariate SN and ST distributions that will be used in defining the joint models with skew distributions considered in this paper. Assume an *m*-dimensional random vector $\boldsymbol{Y}$ follows an *m* variate SN or ST distribution with location vector $\boldsymbol{\mu}$, $m \times m$ positive (diagonal) dispersion matrix $\boldsymbol{\Sigma}$ and $m \times m$ skewness matrix $\boldsymbol{\Delta}(\boldsymbol{\delta}) = diag(\delta_1, \delta_2, \ldots, \delta_m)$ or the degrees of freedom ν, where skewness parameter vector $\boldsymbol{\delta} = (\delta_1, \ldots, \delta_m)^T$. In what follows, we briefly discuss multivariate SN and ST distributions introduced by Sahu et al. (2003), which are suitable for Bayesian inference since they are built using the conditional method. For detailed discussion on properties of SN and ST distributions, see reference (Sahu et al. 2003).

1. Skew-t distribution

An *m*-dimensional random vector $\boldsymbol{Y}$ follows an *m*-variate ST distribution if its probability density function (pdf) is given by

$$f(\boldsymbol{y}|\boldsymbol{\mu}, \sum, \boldsymbol{\Delta}(\boldsymbol{\delta}), \nu) = 2^m t_{m,\nu}(\boldsymbol{y}|\boldsymbol{\mu}, \boldsymbol{A}) P(\boldsymbol{U} > \boldsymbol{0}), \quad (15)$$

where $\boldsymbol{A} = \boldsymbol{\Sigma} + \boldsymbol{\Delta}^2(\boldsymbol{\delta})$, we denote the *m*-variate *t* distribution with parameters $\boldsymbol{\mu}$, $\boldsymbol{A}$ and degrees of freedom ν by $t_{m,\nu}(\boldsymbol{\mu}, \boldsymbol{A})$ and the corresponding pdf by $t_{m,\nu}(\boldsymbol{y}|\boldsymbol{\mu}, \boldsymbol{A})$ henceforth, $\boldsymbol{U}$ follows the *t* distribution $t_{m,\nu+m}(\cdot)$. We denote this distribution by $ST_{m,\nu}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\Delta}(\boldsymbol{\delta}))$. In particular, when $\boldsymbol{\Sigma} = \sigma^2 \boldsymbol{I}_m$ and $\boldsymbol{\Delta}(\boldsymbol{\delta}) = \delta \boldsymbol{I}_m$, equation (15) simplifies to

$$\begin{aligned} f(\boldsymbol{y}|\boldsymbol{\mu}, \sigma^2, \delta, \nu) = 2^m (\sigma^2 + \delta^2)^{-m/2} \frac{\Gamma((\nu+m)/2)}{\Gamma(\nu/2)(\nu\pi)^{m/2}} \left\{ 1 + \frac{(\boldsymbol{y}-\boldsymbol{\mu})^T (\boldsymbol{y}-\boldsymbol{\mu})}{\nu(\sigma^2+\delta^2)} \right\}^{-(\nu+m)/2} \\ \times T_{m,\nu+m} \left[\left\{ \frac{\nu + (\sigma^2+\delta^2)^{-1} (\boldsymbol{y}-\boldsymbol{\mu})^T (\boldsymbol{y}-\boldsymbol{\mu})}{\nu+m} \right\}^{-1/2} \frac{\delta(\boldsymbol{y}-\boldsymbol{\mu})}{\sigma\sqrt{\sigma^2+\delta^2}} \right], \end{aligned} \quad (16)$$

where $T_{m,\nu+m}(\cdot)$ denotes the cumulative distribution function (cdf) of $t_{m,\nu+m}(\mathbf{0}, \boldsymbol{I}_m)$. However, unlike the SN distribution below, the ST density can not be written as the product of univariate ST densities. Here $\boldsymbol{Y}$ are dependent but uncorrelated.

The mean and covariance matrix of the ST distribution $ST_{m,\nu}(\boldsymbol{\mu}, \sigma_2 \boldsymbol{I}_m, \boldsymbol{\Delta}(\boldsymbol{\delta}))$ are given by

$$\begin{aligned} E(\boldsymbol{Y}) &= \boldsymbol{\mu} + (\nu/\pi)^{1/2} \frac{\Gamma((\nu-1)/2)}{\Gamma(\nu/2)} \boldsymbol{\delta}, \\ cov(\boldsymbol{Y}) &= [\sigma^2 \boldsymbol{I}_m + \boldsymbol{\Delta}^2(\boldsymbol{\delta})] \frac{\nu}{\nu-2} - \frac{\nu}{\pi} \left[\frac{\Gamma\{(\nu-1)/2\}}{\Gamma(\nu/2)} \right]^2 \boldsymbol{\Delta}^2(\boldsymbol{\delta}). \end{aligned} \quad (17)$$

It is noted that when $\boldsymbol{\delta} = \mathbf{0}$, the ST distribution reduces to the usual t distribution. In order to have a zero mean vector, we should assume the location parameter $\boldsymbol{\mu} = -(\nu/\pi)^{1/2} \frac{\Gamma((\nu-1)/2)}{\Gamma(\nu/2)} \boldsymbol{\delta}$, which is what we assume in the paper. In order to better understand the shape of an ST distribution, plots of an ST density as a function of the skewness parameter with $\delta = -3, 0, 3$ are shown in Figure 5(a).

By the proposition of Sahu et al. (2003), the ST distribution of $\boldsymbol{Y}$ has a convenient stochastic representation as follows.

$$\boldsymbol{Y} = \boldsymbol{\mu} + \boldsymbol{\Delta}(\boldsymbol{\delta}) |\boldsymbol{X}_0| + \sum^{1/2} \boldsymbol{X}_1, \quad (18)$$

where $\boldsymbol{X}_0$ and $\boldsymbol{X}_1$ are two independent random vectors following $t_{m,\nu}(\mathbf{0}, \boldsymbol{I}_m)$. Note that the expression (18) provides a convenient device for random number generation and for implementation purposes. Let $\boldsymbol{w} = |\boldsymbol{X}_0|$; then $\boldsymbol{w}$ follows an m-dimensional standard t distribution $t_{m,\nu}(\mathbf{0}, \boldsymbol{I}_m)$ truncated in the space $\boldsymbol{w} > \mathbf{0}$ (i.e., the standard half-t distribution). Thus, a hierarchical representation of (18) is given by

$$\boldsymbol{Y} | \boldsymbol{w} \sim t_{m,\nu+m}(\boldsymbol{\mu} + \boldsymbol{\Delta}(\boldsymbol{\delta}) \boldsymbol{w}, \omega \sum), \ \boldsymbol{w} \sim t_{m,\nu}(\mathbf{0}, \boldsymbol{I}_m) \boldsymbol{I}(\boldsymbol{w} > \mathbf{0}), \quad (19)$$

where $\omega = (\nu + \boldsymbol{w}^T \boldsymbol{w})/(\nu + m)$.

2. Skew-normal distribution

We briefly discuss a multivariate SN distribution introduced by Sahu et al. (2003) in this section. An m-dimensional random vector $\boldsymbol{Y}$ follows an m-variate SN distribution, if its pdf is given by

$$f(\boldsymbol{y} | \boldsymbol{\mu}, \sum, \boldsymbol{\Delta}(\boldsymbol{\delta})) = 2^m |\boldsymbol{A}|^{-1/2} \varphi_m \{\boldsymbol{A}^{-1/2}(\boldsymbol{y} - \boldsymbol{\mu})\} P(\boldsymbol{U} > \mathbf{0}), \quad (20)$$

where $\boldsymbol{U} \sim N_m\{\boldsymbol{\Delta}(\boldsymbol{\delta})\boldsymbol{A}^{-1}(\boldsymbol{y} - \boldsymbol{\mu}), \boldsymbol{I}_m - \boldsymbol{\Delta}(\boldsymbol{\delta})\boldsymbol{A}^{-1}\boldsymbol{\Delta}(\boldsymbol{\delta})\}$, and $\phi_m(\cdot)$ is the pdf of $N_m(\mathbf{0}, \boldsymbol{I}_m)$. We denote the above distribution by $SN_m(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\Delta}(\boldsymbol{\delta}))$. An appealing feature of equation (20) is that it gives independent marginal when $\sum = diag(\sigma_1^2, \sigma_2^2, \ldots, \sigma_m^2)$. The pdf (20) thus reduces to

$$f(\boldsymbol{y} | \boldsymbol{\mu}, \sum, \boldsymbol{\Delta}(\boldsymbol{\delta})) = \prod_{i=1}^m \left[\frac{2}{\sqrt{\sigma_i^2 + \delta_i^2}} \varphi \left\{ \frac{y_i - \mu_i}{\sqrt{\sigma_i^2 + \delta_i^2}} \right\} \Phi \left\{ \frac{\delta_i}{\sigma_i} \frac{y_i - \mu_i}{\sqrt{\sigma_i^2 + \delta_i^2}} \right\} \right], \quad (21)$$

where $\phi(\cdot)$ and $\Phi(\cdot)$ are the pdf and cdf of the standard normal distribution, respectively.

The mean and covariance matrix are given by $E(\boldsymbol{Y})=\boldsymbol{\mu}+\sqrt{2/\pi}\boldsymbol{\delta}$, $cov(\boldsymbol{Y}) = \boldsymbol{\Sigma} + (1 - 2/\pi)\boldsymbol{\Delta}^2(\boldsymbol{\delta})$. It is noted that when $\boldsymbol{\delta} = \mathbf{0}$, the SN distribution reduces to usual normal distribution. In addition, the SN distribution is a special case of the ST distribution. That is, the ST distribution reduces to the SN distribution when the degrees of freedom $\nu \to \infty$. In order to have a zero mean vector, we should assume the location parameter $\boldsymbol{\mu}=-\sqrt{2/\pi}\boldsymbol{\delta}$. In order to better understand the shape of an SN distribution, plots of an SN density as a function of the skewness parameter with $\delta = -3$, 0, and 3 are shown in Figure 5(b).

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

$$\boldsymbol{Y}=\boldsymbol{\mu}+\boldsymbol{\Delta}|\boldsymbol{X}_0|+\sum^{1/2}\boldsymbol{X}_1, \quad (22)$$

where $\boldsymbol{X}_0$ and $\boldsymbol{X}_1$ are two independent random vectors with $N_m(\mathbf{0}, \boldsymbol{I}_m)$. Let $\boldsymbol{w} = |\boldsymbol{X}_0|$; then, $\boldsymbol{w}$ follows an m-dimensional standard normal distribution $N_m(\mathbf{0}, \boldsymbol{I}_m)$ truncated in the space $\boldsymbol{w} > \mathbf{0}$. Thus, a two-level hierarchical representation of (22) is given by

$$\boldsymbol{Y}|\boldsymbol{w}\sim N_m(\boldsymbol{\mu}+\boldsymbol{\Delta}\boldsymbol{w}, \sum), \ \boldsymbol{w}\sim N_m(\mathbf{0}, \boldsymbol{I}_m)\boldsymbol{I}(\boldsymbol{w}>\mathbf{0}). \quad (23)$$

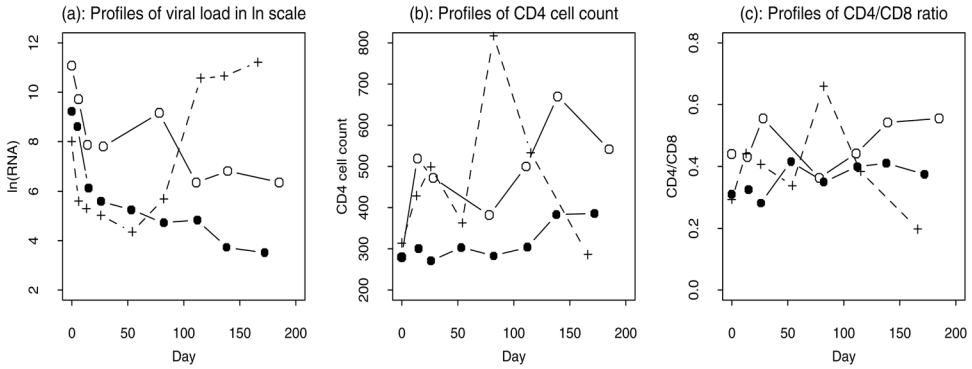

Fig. 1.

Viral load in natural logarithmic scale, CD4 cell count and CD4/CD8 ratio trajectories of three randomly selected patients

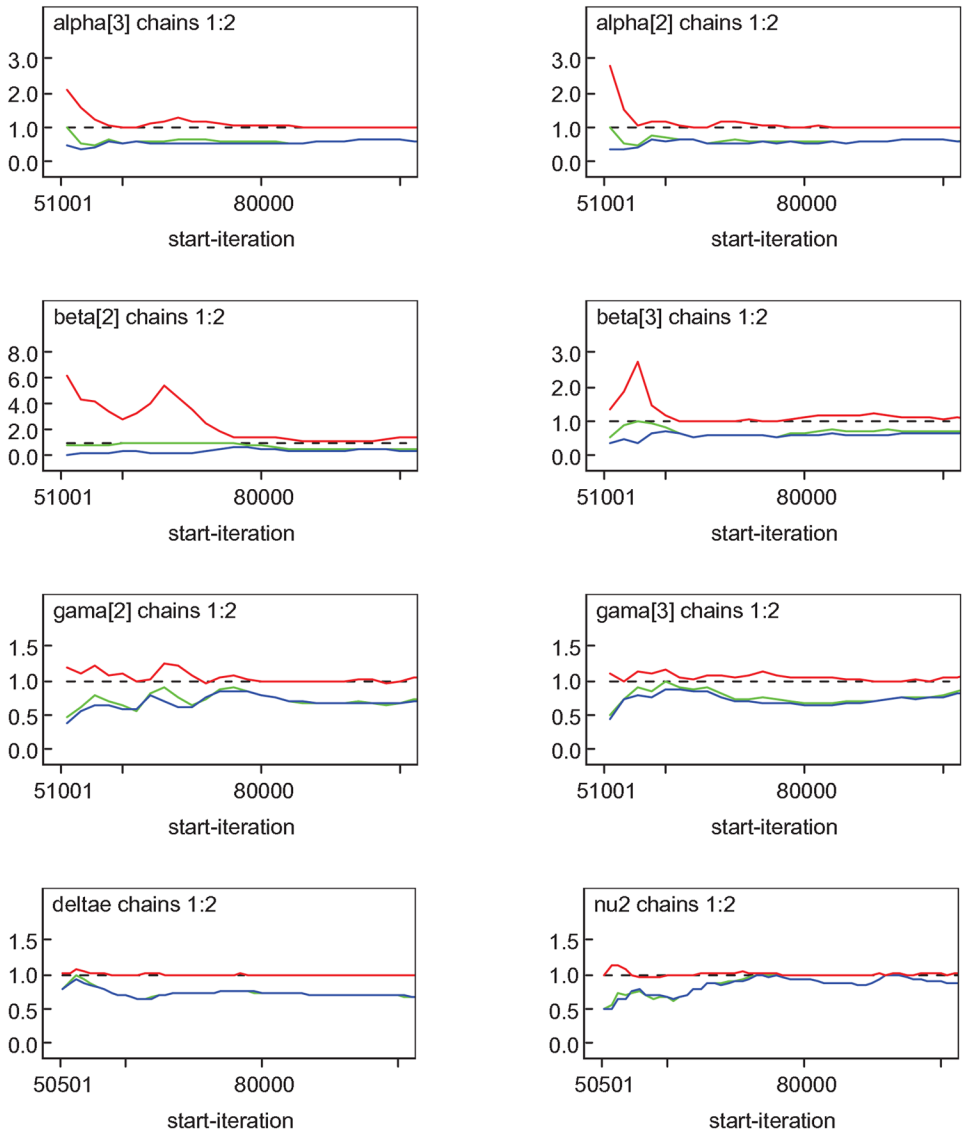

Fig. 2.

Gelman-Rubin (GR) diagnostic plot with the two Markov chains as obtained from the WinBUGS software for the representative parameters. The middle and bottom curves below the dashed horizontal line (indicated by the value one) represent the pooled posterior variance ($\hat{V}$, green color) and average within-sample variance (W, blue color), respectively, and the top curve represents their ratio ($\hat{R}$, red color), which is associated with the potential scale reduction factor.

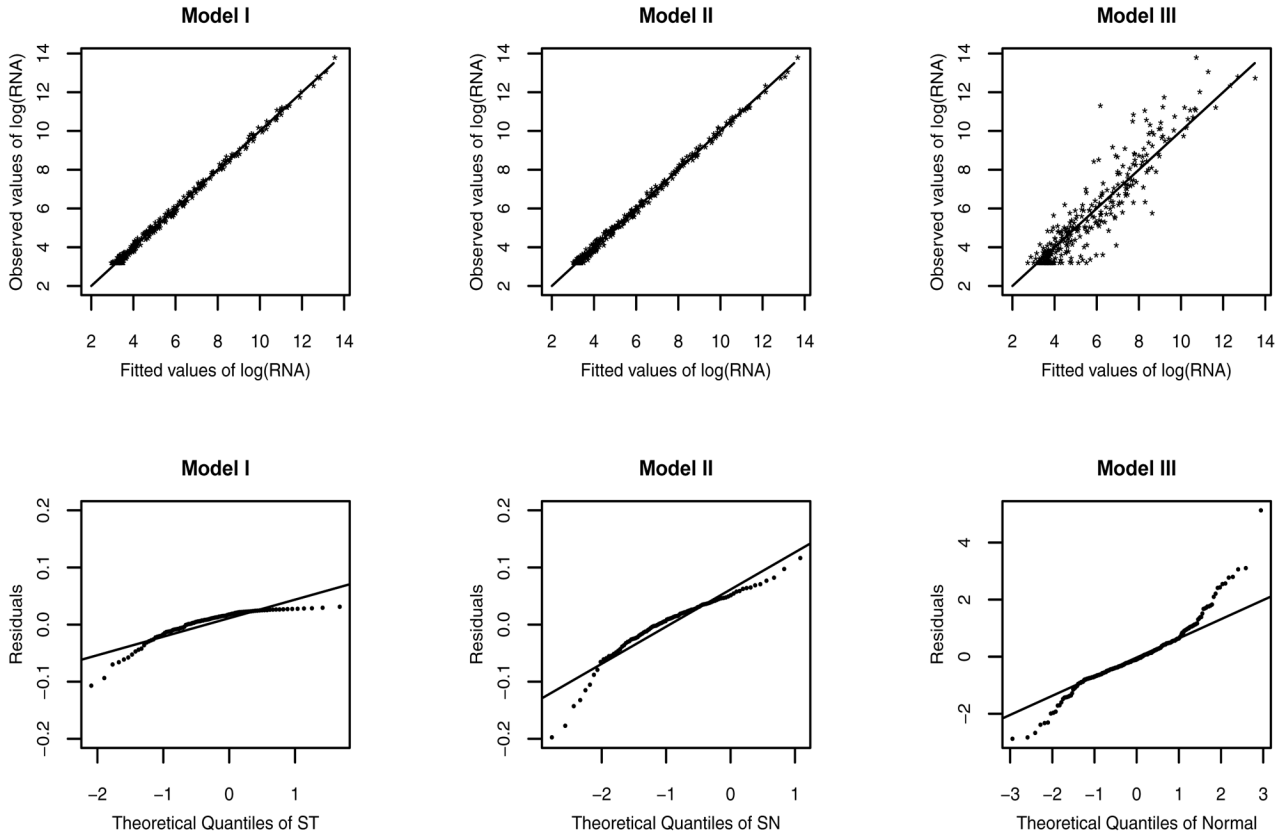

Fig. 3.

The goodness-of-fit: Observed values versus fitted values of ln(RNA) (top panel) and ST, SN or normal Q-Q plot with line (bottom panel)

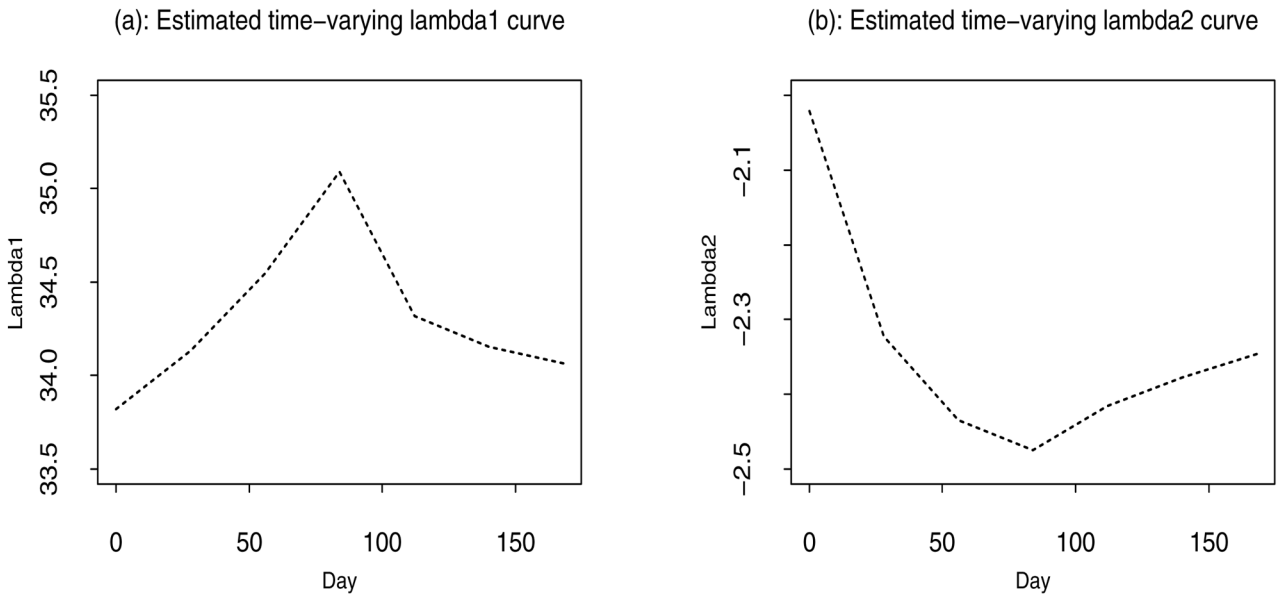

Fig. 4.
The estimated curves of the first and second-phase viral decay rates based on Model I

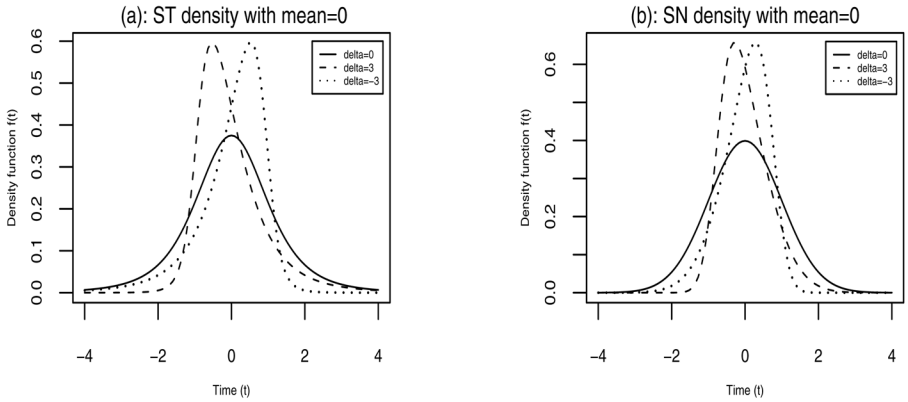

Fig. 5.

The univariate skew-*t* (df $\nu = 4$) and skew-normal density functions with variance $\sigma^2 = 1$ and skewness parameter $\delta = 0, -3$ and 3, respectively.

Table 1

A summary of the estimated posterior mean (PM) of population (fixed-effects) parameters, as well as the corresponding standard deviation (SD) and lower limit (L_{CI}) and upper limit (U_{CI}) of 95% equal-tail credible interval (CI) based on the naive method (NM) and the joint modeling (JM) approach for semiparametric NLME (SJM) model and parametric NLME model(PJM).

Method	Model		β_1	β_2	β_3	β_4	α_1	α_2	α_3	γ_0	γ_1	γ_2	γ_3	γ_4
SJM	I	PM	6.09	33.8	12.2	2.13	−0.10	0.52	−0.14	−2.08	−0.08	0.68	0.03	0.04
		L_{CI}	5.25	26.2	5.29	1.62	−0.41	0.12	−0.59	−2.33	−0.60	−0.60	−0.95	−0.21
		U_{CI}	6.76	40.9	20.1	2.60	0.19	0.95	0.28	−1.86	0.36	2.08	1.15	0.37
		SD	0.38	3.47	3.44	0.27	0.15	0.21	0.23	0.12	0.25	0.67	0.47	0.14
	II	PM	5.23	26.1	10.1	0.41	−0.65	0.58	−0.22	−2.08	−0.08	0.47	0.04	0.01
		L_{CI}	4.63	20.3	3.70	−0.83	−1.17	0.01	−0.76	−2.34	−0.99	−1.32	−0.87	−0.31
		U_{CI}	5.93	32.7	17.7	2.01	0.40	1.09	0.37	−1.83	0.53	1.86	1.17	0.34
		SD	0.33	3.13	3.37	0.81	0.46	0.28	0.29	0.13	0.36	0.71	0.49	0.15
	III	PM	8.35	29.2	7.85	3.70	−0.23	0.64	−0.26	−2.07	−0.06	0.40	−0.01	0.06
		L_{CI}	7.85	22.6	0.66	2.16	−0.48	−0.10	−0.81	−2.30	−0.65	−0.42	−1.32	−0.34
		U_{CI}	8.84	36.9	15.6	5.00	0.03	1.16	0.32	−1.87	0.42	1.30	1.62	0.77
		SD	0.25	3.72	3.81	0.72	0.13	0.27	0.29	0.11	0.27	0.44	0.69	0.24
NM	I	PM	7.87	32.4	10.6	3.94	−0.54	0.62	−0.25	−2.07	−0.07	0.05	0.05	−0.02
		L_{CI}	5.38	26.0	3.43	2.22	−0.96	0.21	−0.68	−2.28	−0.72	−0.68	−1.36	−0.71
		U_{CI}	9.15	40.6	19.2	5.57	0.02	1.02	0.20	−1.87	0.43	1.58	1.40	0.31
		SD	1.23	3.79	4.12	1.00	0.28	0.20	0.22	0.11	0.29	0.56	0.61	0.20
PJM	I	PM	6.13	31.9	13.2	2.28	−0.16	0.57	−0.20	−2.08	−0.06	0.66	0.21	0.03
		L_{CI}	5.55	26.2	7.06	1.21	−0.74	0.18	−0.60	−2.29	−0.67	−0.60	−0.53	−0.50
		U_{CI}	6.70	37.0	19.8	3.03	0.24	0.95	0.24	−1.86	0.47	1.89	1.25	0.30
		SD	0.30	2.84	3.47	0.44	0.26	0.19	0.21	0.11	0.29	0.63	0.46	0.19
SJM	I uniform prior on σ_k	PM	8.50	32.5	12.5	4.62	−0.19	0.70	−0.32	−2.07	−0.23	0.75	0.05	0.04
		L_{CI}	8.10	28.2	6.08	4.21	−0.50	0.33	−0.75	−2.32	−1.05	−0.66	−0.90	−0.34
		U_{CI}	8.89	38.6	17.3	4.99	0.02	1.10	0.11	−1.84	0.53	1.88	1.27	0.59
		SD	0.20	2.63	3.04	0.23	0.13	0.20	0.23	0.12	0.41	0.65	0.48	0.20
SJM	I half-Cauchy prior on σ_k	PM	8.60	35.4	11.7	5.28	−0.17	0.55	−0.18	−2.06	−0.04	0.48	0.12	−0.04
		L_{CI}	7.90	29.8	6.21	5.11	−0.37	0.15	−0.62	−2.29	−0.60	−1.03	−0.84	−0.83

Method	Model		β_1	β_2	β_3	β_4	α_1	α_2	α_3	γ_0	γ_1	γ_2	γ_3	γ_4
		U_{CI}	9.19	42.5	18.7	5.49	0.05	0.98	0.26	−1.85	0.64	1.63	1.29	0.82
		SD	0.33	3.23	3.54	0.10	0.11	0.21	0.22	0.11	0.30	0.65	0.55	0.40

Table 2

A summary of the estimated posterior mean (PM) for the parameters of variance, skewness and degrees of freedom, and corresponding standard deviation (SD) and lower limit (L_{CI}) and upper limit (U_{CI}) of 95% equal-tail credible interval (CI) as well as SSR values based on the naive method (NM) and the joint modeling (JM) approach for semiparametric NLME model (SJM) and parametric NLME model(PJM).

Method	Model		σ_1^2	σ_2^2	δ_ε	δ_e	ν_1	ν_2	SSR
SJM	I	PM	0.05	0.01	0.13	1.17	3.25	3.28	219.8
		L_{CI}	0.03	0.01	−0.12	1.00	3.01	3.01	
		U_{CI}	0.06	0.03	0.33	1.43	3.93	4.07	
		SD	0.01	0.01	0.11	0.08	0.26	0.29	
	II	PM	0.07	0.07	0.28	1.87	–	–	309.3
		L_{CI}	0.03	0.01	−0.40	1.58	–	–	
		U_{CI}	0.13	0.28	0.56	2.12	–	–	
		SD	0.02	0.07	0.29	0.14	–	–	
	III	PM	0.13	1.25	–	–	–	–	410.4
		L_{CI}	0.11	1.03	–	–	–	–	
		U_{CI}	0.15	1.52	–	–	–	–	
		SD	0.01	0.13	–	–	–	–	
NM	I	PM	0.03	0.42	0.15	0.24	3.24	3.45	256.6
		L_{CI}	0.01	0.01	−0.06	−0.28	3.01	3.01	
		U_{CI}	0.06	0.73	0.30	1.27	3.89	4.70	
		SD	0.02	0.24	0.12	0.55	0.24	0.48	
PJM	I	PM	0.02	0.04	0.11	1.10	3.46	3.23	263.5
		L_{CI}	0.01	0.01	−0.12	0.86	3.10	3.01	
		U_{CI}	0.09	0.06	0.24	1.37	4.87	3.86	
		SD	0.03	0.01	0.10	0.12	0.50	0.24	
SJM	I uniform prior on σ_k	PM	0.02	0.01	0.24	1.13	3.23	3.28	213.8
		L_{CI}	0.01	0.01	0.13	0.97	3.10	3.01	
		U_{CI}	0.04	0.07	0.32	1.38	3.87	4.02	
		SD	0.01	0.02	0.05	0.11	0.24	0.27	
SJM	I half-Cauchy prior on σ_k	PM	0.04	0.01	0.19	1.33	3.25	3.17	217.3

Method	Model	σ_1^2	σ_2^2	δ_ε	δ_e	ν_1	ν_2	SSR
	L_{CI}	0.01	0.01	0.01	1.22	3.01	3.01	
	U_{CI}	0.06	0.02	0.23	1.41	3.93	3.58	
	SD	0.01	0.01	0.08	0.05	0.25	0.15	