Statistics and Its Interface Volume 10 (2017) 557–573

Simultaneous Bayesian inference on a finite mixture of mixed-effects Tobit joint models for longitudinal data with multiple features

Yangxin Huang*,†, Jiaqing Chen‡, Ping Yin§, and Huahai Qiu

It often happens in longitudinal studies that some collected data are observed with the following issues. (i) Subjects may not sampled from homogeneous population with a common trajectory; (ii) longitudinal continuous measurements may suffer from a serious departure of normality in which normality assumption may cause lack of robustness and subsequently lead to invalid inference; (iii) some covariates of interest may be difficult to measure accurately due to their nature; and (iv) the response observations may be subject to left-censoring due to a limit of detection (LOD). Inferential procedures will become very complicated when one analyzes data with these features together. In the literature, there has been considerable interest in accommodating heterogeneity, non-normality, LOD or covariate measurement errors in longitudinal data modeling, but, no studies have done concerning all of the four features simultaneously. In this article, simultaneous Bayesian modeling approach based on a finite mixture of nonlinear mixed-effects Tobit joint (NLMETJ) models with skew distributions is developed to study impact of multiple data features together, and to estimate not only model parameters but also class membership probabilities at both population and individual levels. Simulation studies are conducted to assess the performance of the proposed method, and real data example is analyzed to demonstrate the proposed methodologies through comparing potential models with different specifications of error distributions and various scenarios.

AMS 2000 subject classifications: Primary 62F15; secondary 62P10.

Keywords and phrases: Bayesian inference, Left-censoring, Longitudinal data analysis, Measurement errors, Mixture of hierarchical Tobit joint models, Skew distributions.

*Corresponding author.

†The first author's research is partially supported by University South Florida Proposal Enhancement grant (18326).

‡The second author's research is partially supported by the National Natural Science Foundation of China (81671633) and the Natural Science Founds of Hubei Province (2014CFB863).

§The third author's research is partially supported by the National Natural Science Foundation of China (81573262).

1. INTRODUCTION

Modeling of longitudinal data is an active area of biostatistics and statistics research that has received a lot of attention in recent years. A large number of statistical modeling and analysis methods have been suggested for analyzing such data with various features. Researchers may often confront the task of developing inference from samples where longitudinal outcomes for the dependent variable of interest may follow heterogeneous (not homogeneous) characteristics, suffer from a serious departure of normality and be subject to left-censoring due to a limit of detection (LOD), and covariates may often be measured with substantial errors. For example. Modeling AIDS data which will be described in Section 3.1 has many challenges due to the following issues of inherent data features.

First, in the literature, most studies of longitudinal modeling assume that all subjects come from a homogeneous population where large variations between- and within-subjects were accommodated by random-effects and/or time-varying covariates in the models. These typical variations are shown in Figure 1(a), viral load trajectory profiles of six representative patients in AIDS ACTG398 study [10]; these viral load trajectories can be roughly classified into three classes. The detailed data description and class justification are provided in Section 3. We, therefore, can reasonably assume that patients are from a population which consists of three relatively homogeneous classes and, thus, it is motivated to consider a finite mixture of nonlinear mixed-effects (NLME) models for such data set. Second, Figure 1(b) displays the distributions of repeated viral load measurements (in $\log_{10}$ scale) for 379 subjects enrolled in ACTG398 study [10]. It can be seen that, for this data set to be analyzed in this article, the viral load (even after $\log_{10}$ transformation) appears a skewed feature and, thus, a normality assumption may not be quite realistic. Alternatively, an asymmetric distribution such as skew-t (ST) and skew-normal (SN) distributions [1, 25] should be more appropriate than a symmetric (normal) distribution to model the data in longitudinal studies. Third, the response measurements may be subject to left-censoring due to an LOD because of the low sensitivity of current standard assays [24]. It can be seen from Figure 1(a) that for some patients their viral loads are below LOD of 50 copies/mL ($\log_{10}(50) = 1.699$)

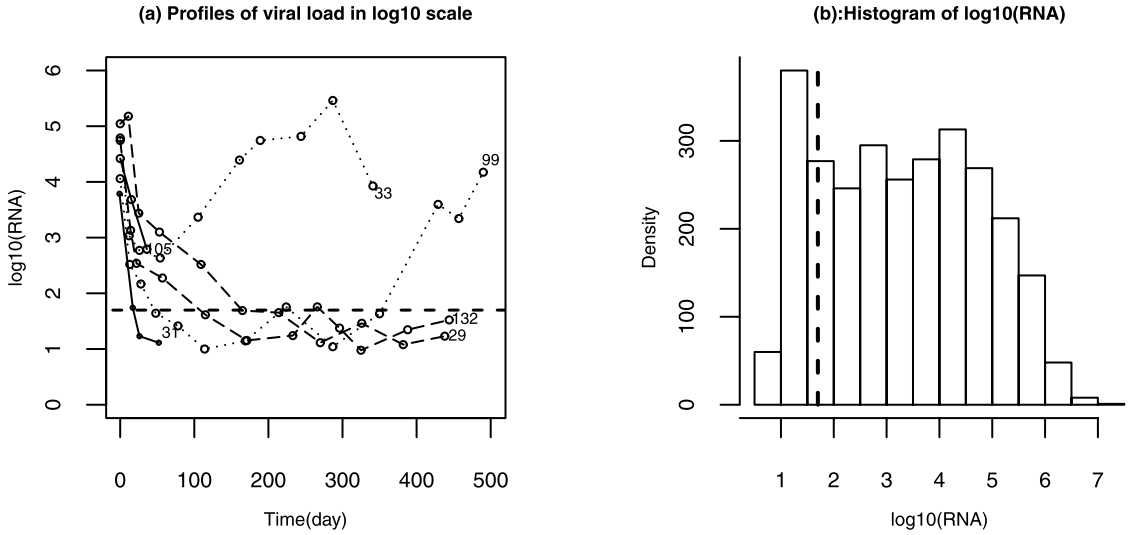

Figure 1. (a) Profile of viral load (response) in log_{10} scale for six representative patients. Trajectory class 1: decrease rapidly and constantly in a short-term period (solid curves); class 2: decrease at the beginning and then maintain stable at a low level (dashed curves) and class 3: decrease at the beginning, but rebound later (dotted curves). (b) Histogram of viral load measured from RNA levels in log_{10} scale in an AIDS clinical trial study. The horizontal and vertical dashed-lines at $1.699 = log_{10}(50)$ are LOD.

which is displayed by the horizontal dashed-line. Last, another feature of a longitudinal data set is the existence of time-varying covariates which suffer from measurement errors. This is usually the case in a longitudinal AIDS study where CD4 cell counts are often measured with substantial measurement errors [15, 18, 31]. Thus, measurement error in covariate should take into account for statistical inference.

Much of the statistical literature on modeling of longitudinal data has focused on the development of models that aim at capturing only specific aspects of motivating case studies [2, 3, 5, 12, 15, 16, 18, 19, 31, 21, 23]. However, to our best knowledge, no studies have been conducted on simultaneously accounting for the biases induced by heterogeneity, non-normality, left-censoring and mismeasured covariate. It is not clear how these features of data may interact and simultaneously influence inferential procedure. Statistical inference and analysis could be complicated dramatically when all of these features arise.

Finite mixture models are used in longitudinal studies [21, 23], where the latent classes corresponding to the mixture components and cluster individuals may provide a better inference. However, most finite mixture models for longitudinal data are currently based on linear (polynomial) [21] or piecewise linear [23] mean functions. The partial reason is that the computation for inference can be conveniently carried out because the likelihood function of a model based on these linear mean functions has a closed form [21]. However, in practice, most longitudinal trajectories appear to be nonlinear patterns.

The goal of this article is to investigate the effects on inference when all of these typical features exist in the longitudinal data. To achieve our objective, this article proposes a finite mixture of NLME Tobit joint (NLMETJ) models with skew distributions, which include ST and SN distributions, to simultaneously account for response with heterogeneity, non-normality, left-censoring and mismeasured covariate. We demonstrate a Bayesian inferential approach to a real data application and conduct simulation studies to estimate both model parameters and class membership probabilities. It is noted that the ST distribution is approximate to the SN distribution when its degrees of freedom approach infinity, and the SN distribution reduces to a normal distribution if skewness parameter is zero. Thus, we use an ST distribution to develop the mixture modeling methodologies, as it can be easily reverted to the normal and SN distributions. In what follows, we consider multivariate ST and SN distributions introduced by [25], which are suitable for a Bayesian inference; see, for example, [15] and [16] for details.

The rest of this paper is organized as follows. Section 2 constructs the finite mixture of NLMETJ models and associated Bayesian inferential method in a general form. In Section 3, we describe an AIDS data set that motivated this research and discuss the specific mixture of joint models, formulated by three mean functions of different mixture components for viral load response. In Section 4, we demonstrate to apply the proposed methodologies to the AIDS data set and report the analysis results. Section 5 is devoted to simulation studies evaluating the performance of the proposed mixture of joint models-based methods. Finally, we conclude the paper with a discussion in Section 6.

2. MIXTURE OF JOINT MODELS AND BAYESIAN INFERENTIAL PROCEDURE

In this section, we present the models and methods in a general form, illustrating that our methods may be applicable in other applications.

2.1 Notation and model framework

Denote the number of subjects by n and the number of measurements on the ith subject by n_i. Let y_{ij} be the value of response for the individual i at time t_{ij} ($i = 1, 2, \ldots, n; j = 1, 2, \ldots, n_i$). In order to introduce Tobit model to deal with observations below LOD in our mixture modeling framework, denote the observed value y_{ij} by (q_{ij}, d_{ij}), where d_{ij} is the censoring indicator and q_{ij} is the latent response variable. The latent q_{ij} is observed, as y_{ij}, if and only if $y_{ij} > \rho$ (a known constant LOD). When q_{ij} is observed we have $d_{ij} = 0$; otherwise we have $d_{ij} = 1$. For simplicity, we consider a single time-varying covariate z_{ij} for the individual i at time t_{ij}. Let the observed data $\Re = \{(\boldsymbol{q}_i, \boldsymbol{d}_i, \boldsymbol{z}_i), i = 1, \ldots, n\}$, where $\boldsymbol{y}_i = (y_{i1}, \ldots, y_{in_i})^T$, $\boldsymbol{q}_i = (q_{i1}, \ldots, q_{in_i})^T$, $\boldsymbol{d}_i = (d_{i1}, \ldots, d_{in_i})^T$ and $\boldsymbol{z}_i = (z_{i1}, \ldots, z_{in_i})^T$.

The various covariate models were investigated in the literature [4, 11, 18, 31]. In the presence of measurement errors in covariate z_{ij}, we adopt a flexible empirical nonparametric mixed-effects model to quantify the mismeasured covariate process as follows.

$$
\begin{aligned}
(1) \qquad & z_{ij} = w(t_{ij}) + h_i(t_{ij}) + \epsilon_{ij} \ (\equiv z^*_{ij} + \epsilon_{ij}), \\
& \boldsymbol{\epsilon}_i \overset{\text{iid}}{\sim} N_{n_i}(\boldsymbol{0}, \sigma_1^2 \boldsymbol{I}_{n_i}),
\end{aligned}
$$

where $w(t_{ij})$ and $h_i(t_{ij})$ are unknown nonparametric smooth fixed-effects and random-effects functions, respectively, and $\boldsymbol{\epsilon}_i = (\epsilon_{i1}, \ldots, \epsilon_{in_i})^T$ follows a multivariate normal distribution. $z^*_{ij} = w(t_{ij}) + h_i(t_{ij})$ are the true (but unobservable) covariate values at time t_{ij}. The fixed smooth function $w(t)$ represents population average of the covariate process, while the random smooth function $h_i(t)$ is introduced to incorporate the large inter-individual variation in the covariate process. We assume that $h_i(t)$ is the realization of a zero-mean stochastic process.

A two-stage modeling for a mixture of NLME model is employed to modulate the response process in connection with covariate. We assume that there are K plausible nonlinear trajectory classes with mean functions $g_k(\cdot)$ ($k = 1, ..., K$), which are known to be specified. The true trajectory mean function of the ith subject might be $g_k(\cdot)$ with unknown probability (population proportion) $\pi_k = P(c_i = k)$ which satisfy $\sum_{k=1}^{K} \pi_k = 1$, where c_i is a latent indicator. For the response process with heterogeneity, skewness and left-censoring due to LOD, a statistical nonlinear trajectory model of individual i, given $c_i = k$, can be formulated by

$$
\begin{aligned}
(2) \qquad & (\boldsymbol{q}_i | c_i = k) = \boldsymbol{g}_k(\boldsymbol{t}_i, \boldsymbol{A}_k\boldsymbol{\beta}_i) + \boldsymbol{e}_i, \\
& \boldsymbol{e}_i \overset{iid}{\sim} ST_{n_i,\nu}(-J(\nu)\delta \boldsymbol{1}_{n_i}, \sigma_2^2 \boldsymbol{I}_{n_i}, \delta \boldsymbol{I}_{n_i}),
\end{aligned}
$$

where $\boldsymbol{t}_i = (t_{i1}, \ldots, t_{in_i})^T$, $\boldsymbol{\beta}_i = (\boldsymbol{\beta}_{i1}, \ldots, \boldsymbol{\beta}_{in_i})$, $\boldsymbol{\beta}_{ij} = (\beta_{1ij}, \ldots, \beta_{sij})^T$ is an $s \times 1$ vector of individual parameters for the ith subject, $J(\nu) = (\nu/\pi)^{1/2} \{\Gamma[(\nu-1)/2]/\Gamma(\nu/2)\}$ with $\Gamma(\cdot)$ being a Gamma function; the vector of random errors $\boldsymbol{e}_i = (e_{i1}, ..., e_{in_i})^T$ follows a multivariate ST distribution with degrees of freedom ν, unknown variance parameter σ_2^2 and skewness parameter δ; $\boldsymbol{g}_k(\boldsymbol{t}_i, \boldsymbol{A}_k\boldsymbol{\beta}_i) = (g_k(t_{i1}, \boldsymbol{A}_k\boldsymbol{\beta}_{i1}), \ldots, g_k(t_{in_i}, \boldsymbol{A}_k\boldsymbol{\beta}_{in_i}))^T$, and $\boldsymbol{A}_k$ ($s \times s$) ($k = 1, \ldots, K$) is known square indicator matrix, of which diagonal elements are either 0 or 1 and off-diagonal elements are all 0. $\boldsymbol{A}_k$ is introduced because the mean functions, specified by the nonlinear functions $g_1(\cdot), \ldots, g_K(\cdot)$, may only involve different subsets of $\boldsymbol{\beta}_i$. By introducing $\boldsymbol{A}_k$, $\boldsymbol{A}_k\boldsymbol{\beta}_i$ will set unrelated elements of $\boldsymbol{\beta}_i$ to 0 in the kth trajectory class. We will illustrate the use of $\boldsymbol{A}_k$ and specify nonlinear mean functions $g_k(\cdot)$ in the real data example below.

Similar to discussion in [19], we can specify model (2) conditionally and marginally as follows.

(3)

$$
(\boldsymbol{q}_i | c_i = k) \sim ST_{n_i,\nu}(\boldsymbol{g}_k(\boldsymbol{t}_i, \boldsymbol{A}_k\boldsymbol{\beta}_i) - J(\nu)\delta \boldsymbol{1}_{n_i}, \sigma_2^2 \boldsymbol{I}_{n_i}, \delta \boldsymbol{I}_{n_i}),
$$

(4)

$$
\boldsymbol{q}_i \sim \sum_{k=1}^{K} \pi_k ST_{n_i,\nu}(\boldsymbol{g}_k(\boldsymbol{t}_i, \boldsymbol{A}_k\boldsymbol{\beta}_i) - J(\nu)\delta \boldsymbol{1}_{n_i}, \sigma_2^2 \boldsymbol{I}_{n_i}, \delta \boldsymbol{I}_{n_i}).
$$

In (4) the vector of mixture probabilities $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_K)^T$ can be also viewed as the mixture weights of all plausible components under framework of the finite mixture models. Model (4) is identifiable, as long as each of the component models is identifiable and distinguishable from each other; when the component models are identifiable but not distinguishable from each other, some constraints may be required to make model (4) identifiable [28].

For individual-specific parameter vector $\boldsymbol{\beta}_{ij}$, we assume,

$$
(5) \qquad \boldsymbol{\beta}_{ij} = \boldsymbol{Z}_{ij}\boldsymbol{\beta} + \boldsymbol{b}_i, \ \boldsymbol{b}_i \overset{iid}{\sim} N_s(\boldsymbol{0}, \boldsymbol{\Sigma_b})
$$

in which $\boldsymbol{\beta} = (\beta_1, ..., \beta_r)^T$ is a vector of universal population parameters; $\boldsymbol{b}_i = (b_{1i}, ..., b_{si})^T$ is a vector of random-effects; $\boldsymbol{\Sigma}_b$ ($s \times s$) is an unknown variance-covariance matrix; $\boldsymbol{Z}_{ij}$ ($s \times r$) is a design matrix including time-independent and/or time-varying covariates such as CD4 cell count. We assume one of the elements in $\boldsymbol{Z}_{ij}$, $\boldsymbol{z}^*_i = (z^*_{i1}, ..., z^*_{in_i})^T$, is a summary of true (but unobservable) time-varying covariate value at time t_{ij}. It is noted that $r \geq s$. Thus, equations (1), (4) and (5) form the finite mixture of NLMETJ models conducted in this paper.

2.2 Simultaneous Bayesian inferential approach

Although a simultaneous inference method based on a joint likelihood for the covariate and response data may

be favorable, the computation associated with the simultaneous likelihood inference in (mixture) joint models for longitudinal data can be intensive and, in particular, may lead to convergence problems; in some cases it can even be computationally infeasible [18, 31]. Here we propose a simultaneous Bayesian inference method via Markov chain Monte Carlo (MCMC) procedure to estimate class membership probabilities and all parameters under framework of a finite mixture of NLMETJ models. Bayesian analysis rests upon computing the conditional posterior distributions of the unknown parameters for inference, given the observed data and weighted by the prior information.

Let $\boldsymbol{\theta} = \{\boldsymbol{\alpha}, \boldsymbol{\beta}, \sigma_1^2, \sigma_2^2, \boldsymbol{\Sigma}_a, \boldsymbol{\Sigma}_b, \delta, \nu\}$ be the collection of unknown parameters in all models except for the mixture weight $\boldsymbol{\pi}$ in (4). Under Bayesian framework, we next need to specify prior distributions for all the unknown parameters as follows.

$$
\begin{aligned}
&\boldsymbol{\alpha} \sim N_p(\boldsymbol{\tau}_1, \boldsymbol{\Gamma}_1),\ \sigma_1^2 \sim IG(\omega_1, \omega_2),\\
&\boldsymbol{\beta} \sim N_r(\boldsymbol{\tau}_2, \boldsymbol{\Gamma}_2),\ \sigma_2^2 \sim IG(\omega_3, \omega_4),\\
&\boldsymbol{\Sigma}_a \sim IW(\boldsymbol{\Omega}_1, \rho_1),\ \boldsymbol{\Sigma}_b \sim IW(\boldsymbol{\Omega}_2, \rho_2),\\
&\delta \sim N(0, \gamma),\ \nu \sim Exp(\nu_0) I(\nu > 2),
\end{aligned} \tag{6}
$$

where the mutually independent Inverse Gamma (IG), Normal (N), Exponential (Exp) and Inverse Wishart (IW) prior distributions are chosen to facilitate computation. The truncation point in exponential prior distribution for ν was chosen to assure a finite variance for ST distribution. The super-parameter matrices $\boldsymbol{\Gamma}_1$, $\boldsymbol{\Gamma}_2$, $\boldsymbol{\Omega}_1$ and $\boldsymbol{\Omega}_2$ can be assumed to be diagonal for convenient implementation. By its definition, the latent indicating variable c_i $(i = 1, ..., n)$ follows a Categorical distribution(Cat)

$$
c_i \overset{iid}{\sim} Cat((1, 2, \ldots, K), (\pi_1, \pi_2, \ldots, \pi_K)), \tag{7}
$$

in which $\boldsymbol{\pi}$ follows a Dirichlet distribution(Dir) [7, 17],

$$
\boldsymbol{\pi} \sim Dir(\eta_1, \eta_2, \ldots, \eta_K). \tag{8}
$$

An MCMC scheme for our mixture model is composed of following two iterative steps:

(i). Sampling class membership indicator c_i, conditional on population parameter $\boldsymbol{\theta}$, and individual random-effects $\boldsymbol{a}_i$ and $\boldsymbol{b}_i$. Generate c_i from

$$
P(c_i = k | \boldsymbol{a}_i, \boldsymbol{b}_i, \boldsymbol{\theta}, \boldsymbol{q}_i) = \frac{\pi_k f(\boldsymbol{q}_i | \boldsymbol{a}_i, \boldsymbol{b}_i, c_i = k, \boldsymbol{\theta})}{\sum_{m=1}^{K} \pi_m f(\boldsymbol{q}_i | \boldsymbol{a}_i, \boldsymbol{b}_i, c_i = m, \boldsymbol{\theta})}, \tag{9}
$$

where $f(\boldsymbol{q}_i | \boldsymbol{a}_i, \boldsymbol{b}_i, c_i = k, \boldsymbol{\theta})$ $(k = 1, ..., K)$ is a conditional density function of $\boldsymbol{q}_i$ based on (3). Then, update the probability $\boldsymbol{\pi}$ for next iteration from distribution

$$
(\boldsymbol{\pi} | num_1, ..., num_K) \sim Dir(\eta_1 + num_1, ..., \eta_K + num_K), \tag{10}
$$

where $num_k = \sum_{i=1}^{n} I(c_i = k), (k = 1, \ldots, K)$, in which $I(\cdot)$ is an indicator function.

(ii). Sampling parameters $\boldsymbol{\theta}$, and individual random-effects $\boldsymbol{a}_i$ and $\boldsymbol{b}_i$, conditional on class membership indicator $\boldsymbol{c} = (c_1, \ldots, c_n)^T$.

Following the study by [25], it can be shown, conditional on c_i determined in step (i), that by introducing the random vector $\boldsymbol{w}_i = (w_{i1}, \ldots, w_{in_i})^T$ and random variable u_i based on the stochastic representation for the ST distribution [12, 15, 16], $\boldsymbol{z}_i$ with random-effects $\boldsymbol{a}_i$, and $\boldsymbol{q}_i$ with random-effects $\boldsymbol{b}_i$ in the presence of left-censoring can be hierarchically formulated as

$$
\begin{aligned}
&\boldsymbol{z}_i | \boldsymbol{a}_i \sim N_{n_i}(\boldsymbol{z}_i^*, \sigma_1^2 \boldsymbol{I}_{n_i}),\ \boldsymbol{a}_i \sim N_q(\boldsymbol{0}, \boldsymbol{\Sigma}_a),\\
&\boldsymbol{q}_i | \boldsymbol{w}_i, u_i, \boldsymbol{b}_i, c_i \sim N_{n_i}(\boldsymbol{g}_{c_i}(\boldsymbol{t}_i, \boldsymbol{A}_{c_i} \boldsymbol{\beta}_i)\\
&+\delta[\boldsymbol{w}_i - J(\nu)\boldsymbol{1}_{n_i}], u_i^{-1} \sigma_2^2 \boldsymbol{I}_{n_i}),\\
&\boldsymbol{w}_i | u_i \sim N_{n_i}(\boldsymbol{0}, u_i^{-1} \boldsymbol{I}_{n_i}) I(\boldsymbol{w}_i > \boldsymbol{0}),\\
&u_i \sim \Gamma(\nu/2, \nu/2),\ \boldsymbol{b}_i \sim N_s(\boldsymbol{0}, \boldsymbol{\Sigma}_{\boldsymbol{b}}),
\end{aligned} \tag{11}
$$

Let $f(\cdot|\cdot)$, $F(\cdot|\cdot)$ and $h(\cdot)$ denote a probability density function (pdf), cumulative density function (cdf) and prior density function, respectively. As we know, *ad hoc* procedures are sometimes used to adjust observations fall below LOD. For instance, a common practice is to impute the censored values by either the LOD or some arbitrary value [30, 31, 15]. Here, instead of arbitrarily imputing the observations below LOD, we use fully Bayesian predictive distributions by incorporating a Tobit model [29] to treat the observations below LOD as left-censoring. That is, Tobit model is introduced to treat the inaccurate measures below LOD as missing values of a latent variable. Conditional on the random variables and some unknown parameters, a detectable measurement $q_{ij} = y_{ij}$ contributes $f(q_{ij} | \boldsymbol{b}_i, w_{ij}, u_i)$, whereas a non-detectable measurement contributes $F(\rho | \boldsymbol{b}_i, w_{ij}, u_i) \equiv P(q_{ij} < \rho | \boldsymbol{b}_i, w_{ij}, u_i)$ in the likelihood. We assume that $\boldsymbol{\alpha}, \boldsymbol{\beta}, \sigma_1^2, \sigma_2^2, \boldsymbol{\Sigma}_a, \boldsymbol{\Sigma}_b, \delta, \nu$ are independent of each other, i.e., $h(\boldsymbol{\theta}) = h(\boldsymbol{\alpha}) h(\boldsymbol{\beta}) h(\sigma_1^2) h(\sigma_2^2) h(\boldsymbol{\Sigma}_a) h(\boldsymbol{\Sigma}_b) h(\delta) h(\nu)$. 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. Thus, the joint posterior density of $\boldsymbol{\theta}$ based on the observed data $\Re = \{(\boldsymbol{q}_i, \boldsymbol{d}_i, \boldsymbol{z}_i)$ and classification indicator $\boldsymbol{c}$ can be given by

$$
f(\boldsymbol{\theta} | \Re, \boldsymbol{c}) \propto \{ \prod_{i=1}^{n} \int \int L_{\boldsymbol{q}_i} f(\boldsymbol{z}_i | \boldsymbol{a}_i) f(\boldsymbol{a}_i) f(\boldsymbol{b}_i) d\boldsymbol{a}_i d\boldsymbol{b}_i \} h(\boldsymbol{\theta}), \tag{12}
$$

where $L_{\boldsymbol{q}_i} = \prod_{j=1}^{n_i} f(q_{ij} | \boldsymbol{b}_i, w_{ij}, u_i)^{1 - d_{ij}} F(\rho | \boldsymbol{b}_i, w_{ij}, u_i)^{d_{ij}} \times f(w_{ij} | w_{ij} > 0) f(u_i)$ is the likelihood for the observed response data, d_{ij} is the censoring indicator defined in Section 2.1.

In general, the integrals in (12) are of high dimension and do not have a 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 and class membership. As an alternative, the MCMC procedure can be used to sample population parameters in $\boldsymbol{\theta}$, and random-effects $\boldsymbol{a}_i$ and $\boldsymbol{b}_i$ from conditional posterior distributions based on (12), using the Gibbs sampler along with the Metropolis-Hastings algorithm. Steps (i) and (ii) are repeated alternatively in iterations of MCMC procedure until convergence is reached. An important advantage of the hierarchical representations in (11) is that they are easily implemented using the freely available WinBUGS software [20]. Note that when WinBUGS software is used to implement our modeling approach, it is not necessary to explicitly specify the full conditional posterior distributions or proportional functions of the density functions of full conditional posterior distributions for parameters to be estimated. Although their derivations are straightforward by working the complete joint posterior density, some cumbersome algebra will be involved. We, thus, omit those here to save space.

3. MOTIVATING DATA SET AND SPECIFICATION OF JOINT MODELS

3.1 Data description and features

The data set that motivated this research is from an ACTG398, which is a randomized, double-blind and placebo-controlled, with an extension to more than 48 week study of the 4-drug class regimen in patients with virologic failure defined by receiving saquinavir, nelfinavir, indinavir, or ritonavir [10]. This study consists of 481 HIV-1 infected patients. The plasma HIV-1 RNA (viral load) is repeatedly quantified at weeks 0, 2, 4, 8, 16, and every 8 weeks until the last patient on study. The number of viral load measurements for each individual varies from 2 to 13. Out of total 481 patients, 379 patients who had more than 2 measurements were included in data analysis. A $\log_{10}$-transformation of viral load was used in the analysis in order to stabilize the measurement variation and to speed up estimation algorithm. The CD4 cell counts were also measured throughout study on a similar scheme. 19.5% of viral load observations were measured below LOD which is 50 copies/mL.

As mentioned in Section 1, the viral load trajectories in ACTG398 study can be roughly classified into three classes with the following consideration. The original motivation based on mixture joint modeling to be conducted in this article was to cluster all patients into two classes with virologic suppression and failure (i.e., viral load rebound), respectively, which is of main interest from a clinical prospective; however, the other class was incorporated (class 1) to capture some patients who withdrew too early to be clustered into either class 2 with virologic suppression or class 3 with viral load rebound. Thus, the number of class components in this analysis was determined empirically based on the viral load trajectory patterns and clinical interpretability. Note that class 1 is referred as a confirmed 'short-term virologic response' due to early dropout and may not indicate virologic suppression in long-term treatment.

As is evident from Figure 1(a), the inter-patient variations in viral load appear to be large and these variations change over time. Previous studies suggest that the inter-patient variations in viral load may be partially explained by time-varying CD4 cell count [15, 18, 31]. CD4 cell counts often have nonnegligible measurement errors, and ignoring these errors may lead to severely misleading results in statistical inference [4].

3.2 Specification of mixture of joint models

With CD4 measures collected over time, we may model the CD4 process to partially address the measurement errors [31]. However, the CD4 trajectories are often complicated, and there is no well established model for the CD4 process. We, thus, model the CD4 process empirically using nonparametric mixed-effects model (1), which works well for such longitudinal data.

The nonparametric mixed-effects model (1) is more flexible than parametric mixed-effects models. To fit model (1), we apply a regression spline method to $w(t)$ and $h_i(t)$. The working principle is briefly described as follows and more details can be found in the literature [5, 6, 33]. 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 $\boldsymbol{\Psi}_p(t) = \{\psi_0(t), \psi_1(t), ..., \psi_{p-1}(t)\}^T$ and $\boldsymbol{\Phi}_q(t) = \{\phi_0(t), \phi_1(t), ..., \phi_{q-1}(t)\}^T$, respectively. That is,

$$
\begin{aligned}
w(t) &\approx w_p(t) = \textstyle\sum_{l=0}^{p-1} \alpha_l \psi_l(t) = \boldsymbol{\Psi}_p(t)^T \boldsymbol{\alpha}, \\
h_i(t) &\approx h_{iq}(t) = \textstyle\sum_{l=0}^{q-1} a_{il} \phi_l(t) = \boldsymbol{\Phi}_q(t)^T \boldsymbol{a}_i,
\end{aligned} \tag{13}
$$

where $\boldsymbol{\alpha} = (\alpha_0, \ldots, \alpha_{p-1})^T$ is a $p \times 1$ vector of fixed-effects and $\boldsymbol{a}_i = (a_{0i}, \ldots, a_{q-1,i})^T$ $(q \leq p)$ is a $q \times 1$ vector of random-effects with $\boldsymbol{a}_i \overset{\text{iid}}{\sim} N_q(\boldsymbol{0}, \boldsymbol{\Sigma}_a)$. Based on the assumption of $h_i(t)$, we can regard $\boldsymbol{a}_i$ as iid realizations of a zero-mean random vector. For our model, we consider natural cubic spline bases with the percentile-based knots.

Substituting $w(t)$ and $h_i(t)$ by their approximations $w_p(t)$ and $h_{iq}(t)$, we can approximate model (1) by the following linear mixed-effects (LME) model.

$$
\begin{aligned}
z_{ij} &\approx \boldsymbol{\Psi}_p(t_{ij})^T \boldsymbol{\alpha} + \boldsymbol{\Phi}_q(t_{ij})^T \boldsymbol{a}_i + \epsilon_{ij} \approx z_{ij}^* + \epsilon_{ij}, \\
\boldsymbol{\epsilon}_i &\overset{\text{iid}}{\sim} N_{n_i}(\boldsymbol{0}, \sigma_1^2 \boldsymbol{I}_{n_i}),
\end{aligned} \tag{14}
$$

Following the studies [5, 18], we use linear combinations of natural cubic splines with percentile-based knots to approximate $w(t)$ and $h_i(t)$ in (1), and set $\psi_0(t) = \phi_0(t) = 1$ and take the same natural cubic splines in the approximations (13) with $q \leq p$ (in order to limit the dimension of random-effects). The values of p and q are determined based on the the Akaike information criterion (AIC) or the Bayesian information criterion (BIC). The AIC/BIC

values are evaluated based on various models (14) with $(p,q) = \{(1,1),(2,1),(2,2),(3,1),(3,2),(3,3)\}$ using CD4 data; it was found that the model with $(p,q) = (3,3)$ has the smallest AIC/BIC values being 1733.5/1749.8. We, thus, incorporate the following nonparametric mixed-effects CD4 covariate model into the Bayesian mixture of NLMETJ models.

$$z_{ij} = (\alpha_0 + a_{0i}) + (\alpha_1 + a_{1i})\psi_1(t_{ij}) + (\alpha_2 + a_{2i})\psi_2(t_{ij}) + \epsilon_{ij}, \tag{15}$$

where z_{ij} is the observed CD4 value at time t_{ij}, $\psi_1(\cdot)$ and $\psi_2(\cdot)$ are two basis functions given in Section 2.1, $\boldsymbol{\alpha} = (\alpha_0, \alpha_1, \alpha_2)^T$ is a vector of population (fixed-effects) parameters, $\boldsymbol{a}_i = (a_{0i}, a_{1i}, a_{2i})^T$ is a vector of random-effects, and $\boldsymbol{\epsilon}_i = (\epsilon_{i1}, \ldots, \epsilon_{in_i})^T \sim N_{n_i}\left(\boldsymbol{0}, \sigma_1^2 \boldsymbol{I}_{n_i}\right)$. In addition, in order to avoid too small or large estimates which may be unstable, we standardize the time-varying covariate CD4 cell counts and rescale the original time (in days) so that the time scale is between 0 and 1.

Viral dynamic models can be formulated through a system of ordinary differential equations (ODE) [24, 30, 22, 13]. Under some reasonable assumptions and simplifications, two useful approximations of ODE solution, which can be used to capture viral load responses, have been proposed as follows.

$$y(t) = \log_{10}(e^{p_1 - \lambda_1 t}), \tag{16}$$

$$y(t) = \log_{10}(e^{p_1 - \lambda_1 t} + e^{p_2 - \lambda_2 t}), \tag{17}$$

where $y(t)$ is the $\log_{10}$ scaled plasma HIV-1 RNA levels at time t. λ_1 and λ_2 are called the first- and second-phase viral decay rates, which may represent the minimum turnover rate of productively infected cells and that of latently or long-lived infected cells, respectively [24]. It is generally assumed that $\lambda_1 > \lambda_2$, which assures that the model is identifiable [30]. The parameters p_1 and p_2 are macro-parameters; e^{p_1} and $e^{p_1} + e^{p_2}$ are the baseline viral load at time $t = 0$ in one- and two-compartment models, respectively. It was noted that both equations (16) and (17) can be only applied to the early segment or longer term of the viral load response with decreasing trajectory patterns (see [14, 19, 32] in detail) and shown in Figure 1(a) (two solid and two dashed curves). It was also noted that for some patients the second-phase viral decay rate, λ_2, may vary over time because they depend on some phenomenological parameters that hide with considerable microscopic complexity and change over time [22]. Negative values of the decay rates may correspond to viral increase and lead to viral rebound [30], suggesting that variation in the dynamic parameters, particularly λ_2, may be partially associated with time-varying covariates such as repeated CD4 cell counts. Thus, it may not be reasonable to assume that second-phase viral decay rate λ_2 is a constant when viral load is rebounded at the later stage during long-term treatment. To model the long-term HIV dynamics, a natural extension of equation (17) is to assume that the second-phase viral decay rate λ_2 changes over time, which may be a function of time-varying covariate such as CD4 cell count to capture the viral load change including viral rebound. Thus, we introduce an extended function as follows.

$$y(t) = \log_{10}(e^{p_1 - \lambda_1 t} + e^{p_2 - \lambda_2(t)t}), \tag{18}$$

Based on discussion above, we consider one- and two-compartment models with constant decay rate(s) for trajectory classes 1 and 2 defined in Section 1, respectively, and a two-compartment model with a time-varying decay rate in the second compartment for trajectory class 3. Thus, the mean functions of $K = 3$ components in the mixture model are specified by

- One-compartment model with a constant decay rate for class 1 trajectory
$$g_1(t_{ij}, \boldsymbol{A}_1\boldsymbol{\beta}_{ij}) = \log_{10}(e^{p_{1i} - \lambda_{1i} t_{ij}}), \tag{19}$$
- Two-compartment model with constant decay rates for class 2 trajectory
$$g_2(t_{ij}, \boldsymbol{A}_2\boldsymbol{\beta}_{ij}) = \log_{10}(e^{p_{1i} - \lambda_{1i} t_{ij}} + e^{p_{2i} - \lambda_{2i} t_{ij}}), \tag{20}$$
- Two-compartment model with constant and time-varying decay rates for class 3 trajectory
$$g_3(t_{ij}, \boldsymbol{A}_3\boldsymbol{\beta}_{ij}) = \log_{10}(e^{p_{1i} - \lambda_{1i} t_{ij}} + e^{p_{2i} - \lambda^*_{2ij} t_{ij}}). \tag{21}$$

In (19)–(21),

$$\begin{aligned}
\beta_{1i} = p_{1i} &= \beta_1 + b_{1i},\ \beta_{2i} = \lambda_{1i} = \beta_2 + \beta_3 z_{i0} + b_{2i},\\
\beta_{3i} = p_{2i} &= \beta_4 + b_{3i},\ \beta_{4i} = \lambda_{2i} = \beta_5 + b_{4i},\\
\beta_{5ij} = \lambda^*_{2ij} &= \beta_5 + \beta_6 z^*_{ij} + b_{4i},\\
\boldsymbol{\beta}_{ij} &= (\beta_{1i}, \beta_{2i}, \beta_{3i}, \beta_{4i}, \beta_{5ij})^T,\\
\boldsymbol{\beta} &= (\beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6)^T,\\
\boldsymbol{A}_1 &= \mathrm{diag}(1,1,0,0,0),\\
\boldsymbol{A}_2 &= \mathrm{diag}(1,1,1,1,0),\\
\boldsymbol{A}_3 &= \mathrm{diag}(1,1,1,0,1),
\end{aligned}$$

where z_{i0} is the baseline CD4 and z^*_{ij} is true (but unobservable) value of CD4 at time t_{ij} defined in (15). The decay rate of the second compartment in (21), β_{5ij}, is time-varying due to z^*_{ij}, but other parameters in $\boldsymbol{\beta}_{ij}$ are time independent. As mentioned previously, the mean functions in different components may involve different subsets of $\boldsymbol{\beta}_{ij}$; for example, $g_1(\cdot)$ only involves parameters β_{1i} and β_{2i}, and $g_2(\cdot)$ and $g_3(\cdot)$ share the same parameters β_{1i}, β_{2i}, and β_{3i} but have different second-phase decay rate, β_{4i} and β_{5ij}, respectively. The diagonal indicator matrices, $\boldsymbol{A}_1$, $\boldsymbol{A}_2$, and $\boldsymbol{A}_3$, determine which elements of $\boldsymbol{\beta}_{ij}$ are involved and set other unrelated parameters to be 0 in the mean functions $g_1(\cdot)$, $g_2(\cdot)$, and $g_3(\cdot)$, respectively. It is noted that (21) is a natural extension of (20) to consider a time-varying decay rate for capturing

viral rebound in class 3 trajectories. With this mixture clustering, our mixture modeling can be used to estimate probabilities of class membership which is either viral rebound eventually, suggesting virologic failure (class 3) or viral decrease continuously, indicating a confirmed short-term virologic response (class 1) and a confirmed long-term virologic response (class 2).

4. ANALYSIS OF AIDS CLINICAL DATA

4.1 Model implementation

As shown in Figure 1(b), the histogram of viral load in $\log_{10}$ scale clearly indicates the asymmetric feature. Thus, it seems plausible to fit models with a skew distribution to the data. With this information, based on the finite mixture of NLMETJ models the following four statistical models with specifying different distributions for viral load response in the presence of left-censoring and the CD4 covariate model (15) were employed to compare their performance:

- **Model N:** A mixture of NLMETJ model where the three mean functions specified by (19)–(21) with the normal distribution for response model error.
- **Model SN:** A mixture of NLMETJ model where the three mean functions specified by (19)–(21) with the SN distribution for response model error.
- **Model ST:** A mixture of NLMETJ model where the three mean functions specified by (19)–(21) with the ST distribution for response model error.
- **Model NL:** A commonly used NLMETJ model where the mean function specified by (21) alone with the SN distribution for response model error.

We conducted the following three scenarios. First, because a normal distribution is a special case of an SN distribution when the skewness parameter is zero, while the ST distribution reduces to the SN distribution when degrees of freedom are large, we investigated how asymmetric distributions for model error (Models ST and SN) contribute to modeling results in comparison with a symmetric (normal) distribution for model error (Model N). Second, we estimated the model parameters by using the 'naive' method, which ignores measurement error in CD4 covariate. That is, the 'naive' method uses only the observed CD4 values z_{ij} rather than unobservable CD4 values z_{ij}^* in equation (21). We used it as a comparison to the joint modeling approach proposed in this article to investigate how the measurement errors in CD4 contribute to modeling results. Finally, we further compared Model NL (ignoring data feature of heterogeneous population) with Model SN to explore how heterogeneous feature influences modeling results.

To carry out the Bayesian inference, we took weakly-informative prior distributions for the parameters in Models N, SN and ST. In particular, (i) fixed-effects were taken to be independent normal distribution $N(0, 100)$ for each element of the population parameter vectors $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$; (ii) we assume a noninformative inverse Gamma prior distribution $IG(0.01, 0.01)$, which has mean 1 and variance 100, for variance parameters σ_1^2 and σ_2^2; (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 $IW(\boldsymbol{\Omega}_1, \upsilon_1)$ and $IW(\boldsymbol{\Omega}_2, \upsilon_2)$, where the diagonal elements for diagonal variance matrix $\boldsymbol{\Omega}_1$ and $\boldsymbol{\Omega}_2$ were 0.01, and $\upsilon_1 = \upsilon_2 = 4$; (iv) the degrees of freedom parameter ν followed truncated exponential distribution $\nu \sim Exp(0.1)I(\nu > 2)$; (v) for the skewness parameter δ, we chose independent normal distributions $N(0, 100)$; and (vi) we set hyper-parameters of Dirichlet distribution in (8), $\eta_1 = \eta_2 = \eta_3 = 1$, assuming individuals have equal probabilities of coming from any one of three classes initially.

The MCMC sampler was implemented using WinBUGS software [20] interacted with a function called bugs in a package R2WinBUGS of R, and the program code is available from authors upon request. When the MCMC procedure was applied to the actual clinical data, convergence of the generated samples was assessed using standard tools within WinBUGS software such as Gelman-Rubin diagnostics [9]. Figure 2 shows the dynamic version of Gelman-Rubin diagnostics based on Model SN as obtained from the WinBUGS software for the representative parameters where the three curves are given: the middle and bottom curves below the horizontal dashed-line (indicated by the value one) represent the pooled posterior variance ($\hat{V}$) and average within-sample variance (W), respectively, and the top curve represents their ratio ($\hat{R}$). 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 convergence diagnostics observed, we proposed that, after an initial number of 50,000 burn-in iterations of three chains of length 100,000, every 50th MCMC sample was retained from the next 50,000 for each chain. Thus, we obtained a total of 3,000 samples of targeted posterior distributions of the unknown parameters for statistical inference.

4.2 Comparison of modeling results

Bayesian modeling approach in conjunction with the mixture of NLMETJ models with different specification of error distributions was used to fit the viral load and CD4 data with multiple features simultaneously. Tables 1 and 2 present the population posterior mean (PM), the corresponding standard deviation (SD) and 95% credible interval (CI) for fixed-effects parameters based on the proposed models with two methods. The following findings are obtained for the results of estimated parameters.

In the mixture of response models, the findings, particularly for the fixed-effects (β_5, β_6), which are parameters related to second-phase viral decay rate, show that these estimates are different from zero for Models N, SN and ST since the 95% credible intervals do not contain zero. Nevertheless, for the estimate of the coefficient of CD4 covariate β_6, its estimate is significantly positive. This means that CD4 has a significantly positive effect on the second-phase

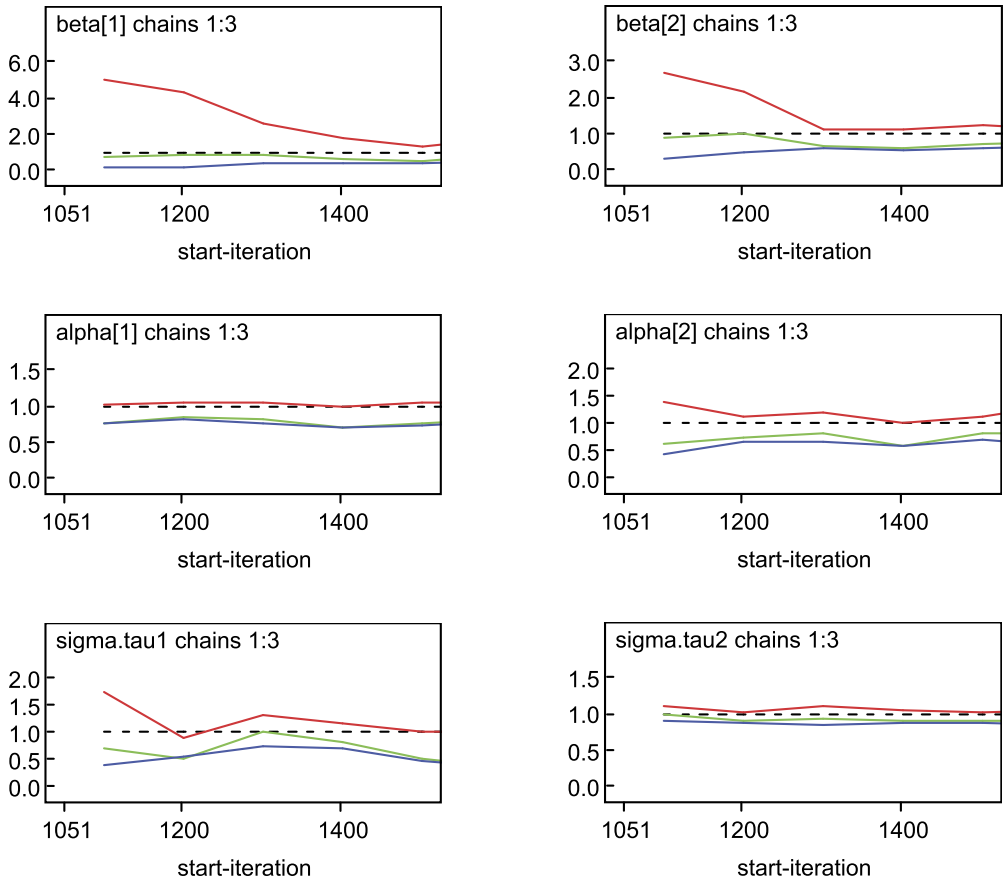

Figure 2. Gelman-Rubin (GR) diagnostic plot based on the NLME mixture model with three Markov chains as obtained from the WinBUGS software for representative parameters. The middle and bottom curves below the horizontal dashed-line (indicated by the value one) represent the pooled posterior variance ($\hat{V}$) and average within-sample variance ($\hat{W}$), respectively, and the top curve above the horizontal dashed-line represents their ratio ($\hat{R}$).

viral decay rate, suggesting that the CD4 covariate may be an important predictor of the second-phase viral decay rate during the treatment. The fixed-effects (β_2, β_3), which are parameters of the first-phase viral decay rate, show that β_2 is significantly estimated, while the estimate of β_3, the coefficient of baseline CD4 count, is significantly positive, indicating that the baseline CD4 has positive effect on the first-phase viral decay rate. In addition, there are marked difference in posterior mean of the variance (scale parameter) σ_2^2 (0.37 vs. 0.01) in comparison of (symmetric) Model N with (asymmetric) Models SN and ST. The estimated values of the variance for both Models SN and ST are much smaller than that of Model N because the former models take into account skewness of the data while the latter does not. The estimates of the skewness parameter (δ) of Models SN and ST are, respectively, 0.64 with 95% CI (0.27, 0.86) and 0.44 with 95% CI (0.11, 0.68). This finding suggests that there is a significantly positive skewness in the data and confirms the fact that the distribution of the original data is skewed even after taking $\log_{10}$-transformation (see Figure 1(b)). Thus, incorporating a skewness parameter in modeling skewed data is recommended. Furthermore, the estimate (0.44) of the skewness parameter based on Model ST is less than that (0.64) based on Model SN. This may be due to the fact that an additional parameter ν (degrees of freedom) for heaviness in the tails was estimated to be 3.80 trading-off the effect of skewness. For parameter estimates of the CD4 covariate model (15), the estimates (-0.03, -0.19 and -0.21) of the intercept α_1 are negatively different from zero for all of the three models. The estimates (0.89 and 0.91) of the linear coefficient α_2 based on Models SN and ST are around six times larger than that (0.15) based on Model N, indicating that the estimated CD4 trajectories appear positive linear patterns for all of the three models. The estimates (26.7, 6.60 and -1.69) of the quadratic coefficient α_3 are quite different and, in particular, their estimates may not be statistically significant for Models SN and ST since the 95% credible intervals contain zero. The estimates (0.78, 0.76 and 0.78) of the variance σ_1^2 are almost equivalent for the three models.

Figure 3 displays the six representative individual fitting curves of viral load observations using the joint modeling ap-

Table 1. Summary of estimated posterior mean (PM) of population (fixed-effects) parameters, the corresponding standard deviation (SD), and lower limit (L_{CI}) and upper limit (U_{CI}) of 95% equal-tail credible interval (CI) based on 'naive' method (NM) and joint modeling approach (JM) for a finite mixture of NLMETJ models (Models N, SN and ST) and a commonly used NLMETJ model (Model NL)

Method	Model		β_1	β_2	β_3	β_4	β_5	β_6	α_1	α_2	α_3
JM	N	PM	10.6	19.0	24.6	6.67	-2.85	40.6	-0.03	0.15	26.7
		L_{CI}	10.3	0.65	16.1	6.23	-4.87	31.2	-0.08	0.07	15.5
		U_{CI}	10.8	39.3	31.9	7.07	-0.85	53.4	-0.01	0.38	38.5
		SD	0.12	10.4	4.12	0.22	1.02	5.70	0.04	0.13	5.84
JM	SN	PM	10.3	26.0	18.8	6.30	-3.79	12.5	-0.19	0.89	6.60
		L_{CI}	10.1	11.5	11.7	5.60	-4.83	8.22	-0.24	0.66	-8.23
		U_{CI}	10.6	44.9	24.0	6.76	-2.96	15.4	-0.13	1.13	21.7
		SD	0.15	9.80	2.59	0.44	0.54	2.13	0.04	0.17	8.19
JM	ST	PM	10.2	25.7	25.3	6.80	-2.64	10.9	-0.21	0.91	-1.69
		L_{CI}	9.98	2.30	17.9	6.53	-3.66	5.41	-0.27	0.75	-13.7
		U_{CI}	10.4	44.2	36.2	7.21	-1.84	15.1	-0.16	1.03	7.46
		SD	0.13	12.1	5.69	0.20	0.55	2.80	0.03	0.06	5.82
JM	NL	PM	10.5	31.3	19.7	6.11	-3.85	10.3	-0.18	0.74	6.57
		L_{CI}	10.3	8.89	14.3	5.73	-5.32	7.23	-0.22	0.65	-1.81
		U_{CI}	10.7	50.7	24.6	6.78	-2.53	13.3	-0.15	0.88	14.6
		SD	0.12	12.4	3.00	0.42	0.88	1.61	0.02	0.06	4.27
NM	SN	PM	10.3	44.0	16.0	6.45	-1.92	2.27	–	–	–
		L_{CI}	9.82	22.5	6.38	6.00	-2.40	1.07	–	–	–
		U_{CI}	10.7	65.9	24.3	6.75	-1.25	3.07	–	–	–
		SD	0.28	15.8	5.82	0.21	0.37	0.69	–	–	–

proach based on Models N, SN and ST. The following findings are observed from joint modeling results. (i) In general, Models N, SN and ST provided a reasonably good fit to the observed data above LOD for most patients in this study, although the fitting for a few patients was not completely satisfactory due to unusual viral load fluctuation patterns for these patients. (ii) In particular, the estimated individual trajectories for Models SN and ST fit the originally observed values above LOD more closely than those for Model N. Note that the lack of smoothness in Models SN and ST estimates of individual trajectories is understandable since a random component $\boldsymbol{w}_i$ was incorporated in the expected function (see (11) for details) according to the stochastic representation feature of the skew distributions for "chasing the data" to this extent. The predicted values where the viral load observations are below LOD appear quite different for the three models (see detailed discussion below).

We now focus on the lower end of the distribution of the viral load where there is left-censoring due to LOD, which is $\log_{10}(50) = 1.699$ here. In our analysis, we treated those inaccurate viral loads below LOD as missing values and predict them by introducing the Tobit model into the proposed mixture models. Note that the main advantage of the Tobit model is that it deals correctly with observations below LOD whether we want predict them or not based on a latent variable approach. The fitted results of these models for values below LOD are depicted in Figure 4, where the histograms show the observed distributions but inaccurate values (Figure 4(a)) below LOD and the predicted values below LOD under Models N, SN and ST (Figure 4(b-d)), where the vertical dashed-lines represent the LOD value at 1.699. It can be seen from the histograms that most observed but inaccurate values are piled up at the range (0.7, 1.7) in the upper-left histogram, whereas the predicted values of the unobserved viral load below LOD for Models N, SN and ST are spread out as expected. However, for Model N some predicted values exceeded the LOD, suggesting bad fits, while both Models SN and ST show an improvement over Model N by giving no predicted values greater than LOD. When we compare Models SN and ST in terms of their performance in predicting viral loads below LOD, we see that Model SN gives a little more plausible values in the sense that the distribution of the predicted viral loads is portrayed as proportionally increasing towards the LOD. This distribution is relatively smooth and closely fits the lower part of the whole distribution of the predicted viral load values based on Model SN, implying that Model SN is the best model. This finding also confirms the conclusions made using other criteria (see below).

Table 2. Summary of estimated posterior mean (PM) for parameters of scale, skewness, and degree 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 DIC, EPD, SSR values and data features based on 'naive' method (NM) and joint modeling approach (JM) for a finite mixture of NLMETJ models (Models N, SN and ST) and a commonly used NLMETJ model (Model NL)

Method	Model		σ_1^2	σ_2^2	δ	ν	DIC	EPD	SSR	Data features
JM	N	PM	0.78	0.37	–	–	14368.9	0.702	1028	heterogeneity
		L_{CI}	0.74	0.34	–	–				left-censoring
		U_{CI}	0.84	0.40	–	–				mismeasured covariate
		SD	0.03	0.02	–	–				
JM	SN	PM	0.76	0.01	0.64	–	11572.8	0.006	7.783	heterogeneity
		L_{CI}	0.72	0.01	0.27	–				left-censoring
		U_{CI}	0.81	0.02	0.86	–				mismeasured covariate
		SD	0.02	0.01	0.09	–				skewness
JM	ST	PM	0.78	0.01	0.44	3.80	12099.1	0.017	23.27	heterogeneity
		L_{CI}	0.73	0.01	0.11	3.12				left-censoring
		U_{CI}	0.83	0.02	0.68	4.77				mismeasured covariate
		SD	0.03	0.01	0.03	0.43				skewness
JM	NL	PM	0.79	0.01	0.52	–	13031.1	0.005	7.450	left-censoring
		L_{CI}	0.74	0.01	0.21	–				mismeasured covariate
		U_{CI}	0.85	0.02	0.73	–				skewness
		SD	0.03	0.01	0.06	–				
NM	SN	PM	–	0.01	0.43	–	12056.3	0.353	499.1	heterogeneity
		L_{CI}	–	0.01	0.27	–				left-censoring
		U_{CI}	–	0.02	0.72	–				skewness
		SD	–	0.01	0.07	–				

To assess the goodness-of-fit for the proposed statistical models, Figure 5 presents the N, SN, and ST Q-Q plots based on Models N, SN and ST. It can be seen that Models SN and ST provided better fit to observed data compared with Model N. Further, Model SN did an even better job than Model ST in accounting for skewness. This finding is confirmed by their sums of squared residuals (SSR), formulated by SSR $= \sum_{i,j} \left(y_{fitted,ij} - y_{obs,ij}\right)^2$, which are 1028 (Model N), 7.783 (Model SN), and 23.37 (Model ST), respectively.

To select the best model that fits the data adequately, a Bayesian selection criterion, known as deviance information criterion (DIC) suggested by [26], 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 [8] in detail). This criterion chooses the model where the discrepancy between predictive values and observed values is the lowest. For mixture models, the structure of DIC does not allow for automatic computation in WinBUGS software. We wrote R code to calculate estimated DIC values, which are 14368.9 (Model N), 11572.8 (Model SN) and 12099.1 (Model ST), respectively. Thus, based on the DIC, the results indicate that Model SN is the best fitting model and Model ST is next, supporting the contention of a departure from normality. This finding is confirmed by the results of the EPD values (see Table 2). These results are also consistent with those in diagnosis of the goodness-of-fit displayed in Figure 5, indicating that Model SN performs best. In summary, our results may suggest that it is important to assume a skew distribution in the mixture models for the viral load response in order to achieve reliable results, in particular, if the data exhibit skewness. Based on these findings, we further report our results in detail only for the best Model SN below.

4.3 Analysis results based on Model SN

As mentioned in Section 1, one of the primary objectives in this data analysis was to cluster all individuals into 3 classes of viral load trajectories: (i) decrease rapidly and constantly in a short-term period, (ii) decrease at the beginning and then maintain stable at a low level, and (iii) decrease at the beginning, but rebound later. Based on the mixture modeling, we are able to obtain a summary of class

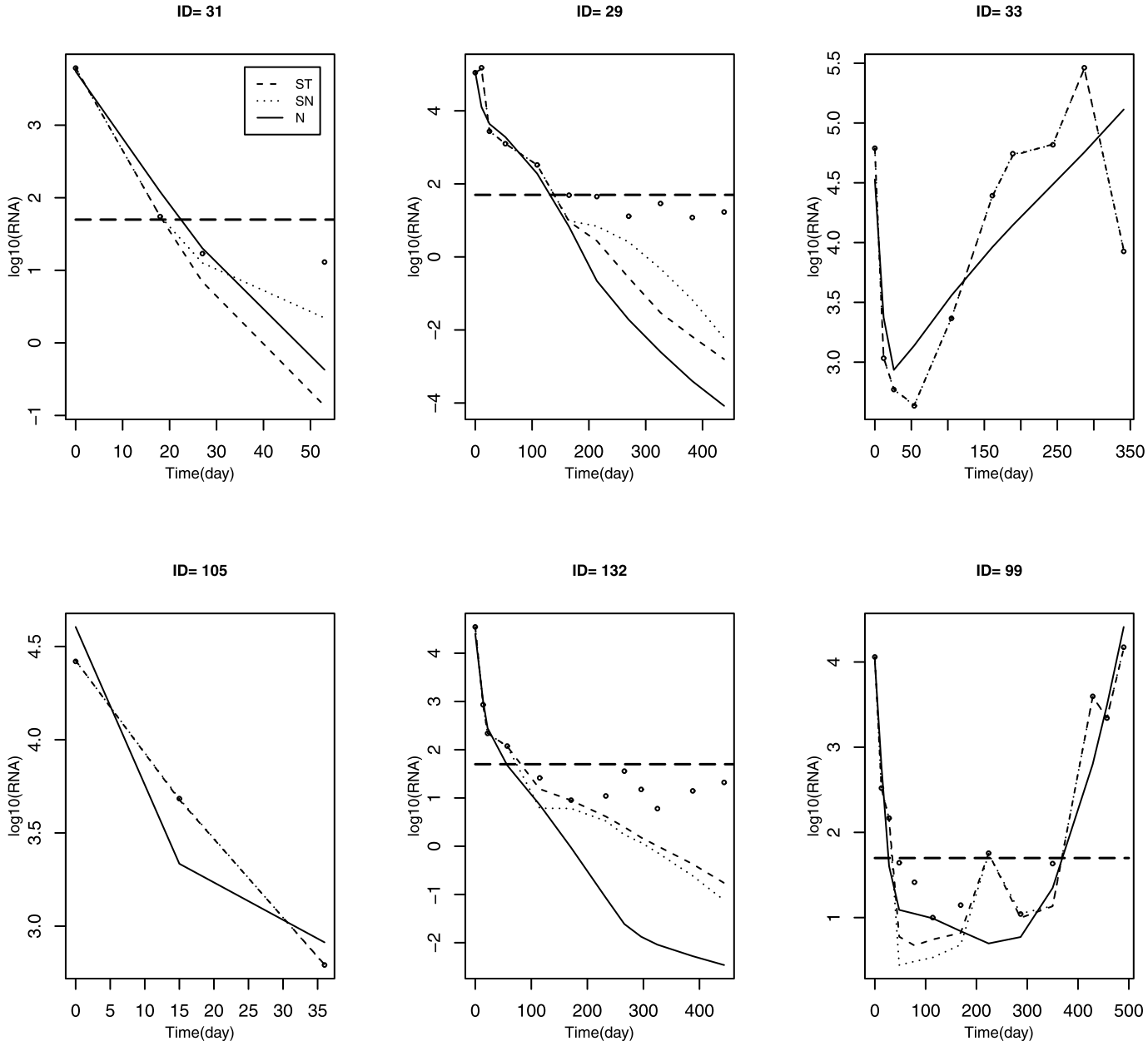

Figure 3. Individual fitted curves of viral load for six representative patients based on Models N (solid curve), SN (dotted curve) and ST (dashed curve). Patients 31 and 105 are from class 1 with probabilities 97% *and* 88%*; patients 29 and 132 belong to class 2 with probabilities* 95% *and* 100%*; and patients 33 and 99 are from class 3 with both probabilities being* 100%*. The observed values are indicated by circle "∘". The horizontal dashed-lines represent LOD at 1.699* $= log_{10}(50)$.

membership at both the population and individual levels. At population level, the MCMC procedure yields samples from the posterior distribution of $\boldsymbol{\pi} = (\pi_1, \pi_2, \pi_3)^T$ in (10), the population proportion of individuals in each class. The estimates of population proportion and associated 95% CI of (π_1, π_2, π_3) for three classes are 25.35% (23.26%, 27.53%), 36.60% (33.87%, 39.41%) and 38.05% (35.19%, 40.88%), respectively. Thus, out of 379 patients, the patterns of changing viral load of 96, 139 and 144 patients followed classes 1, 2 and 3, respectively. It can be seen that classes 1 and 2 (decrease, and decrease and maintain stable) have a cumulated proportion with 61.95%, and class 3 has proportion of 38.05% (decrease and rebound later). This indicates that a confirmed virologic responses were observed in 61.95% of the patients (in classes (i) and (ii)). At individual level, the posterior probability of individual i belonging to the kth $(k = 1, 2, 3)$ class, $p_{ik} = E[I(c_i = k)]$, can be approximated by $M^{-1} \sum_{l=1}^{M} I(c_i^{(l)} = k)$, in which $c_i^{(l)}$ is class member-

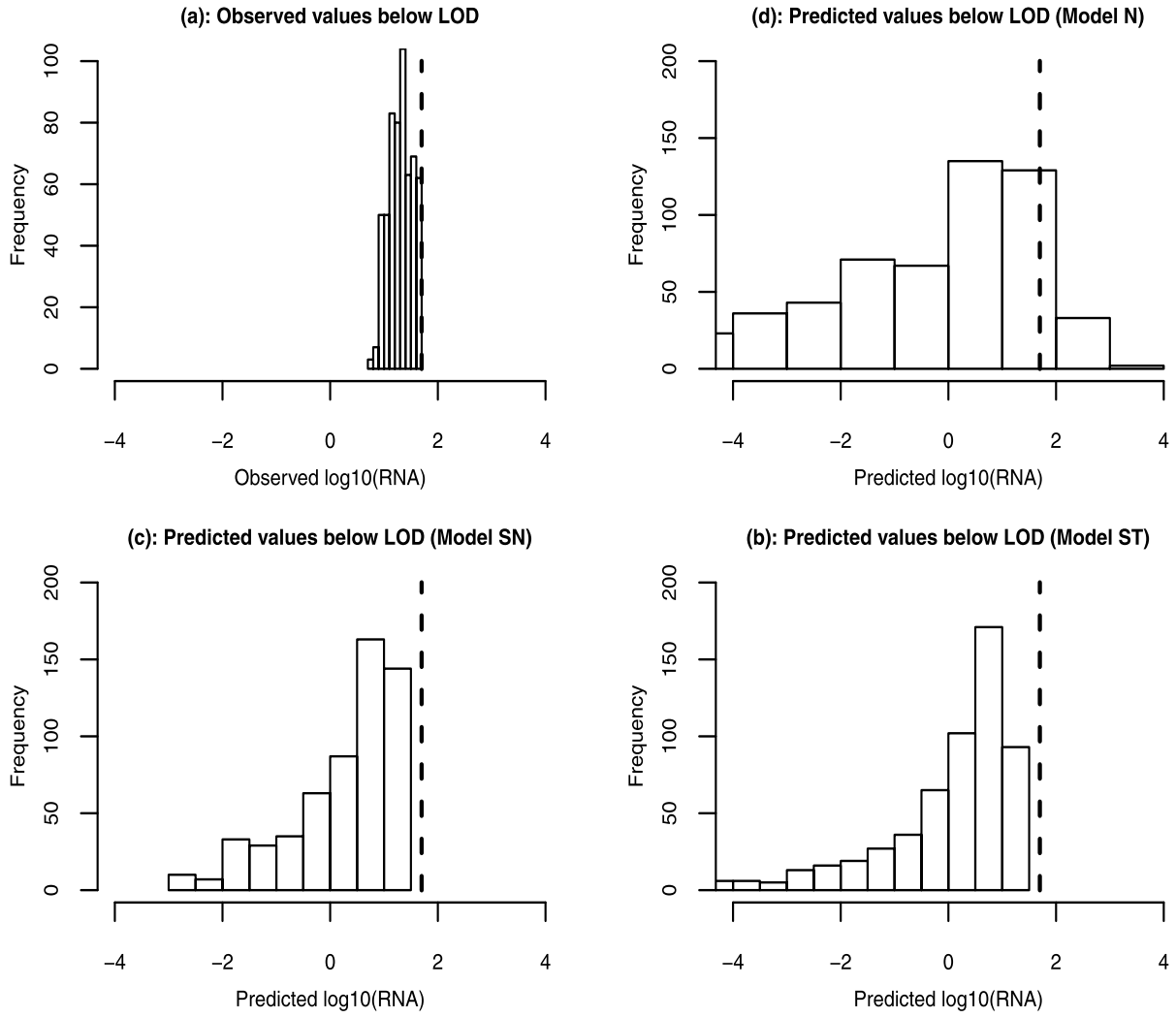

Figure 4. Histograms of (a) inaccurate raw data below LOD, and predicted values of viral load below LOD based on (b) Model N, (c) Model SN and (d) Model ST. The vertical dashed-lines represent LOD at 1.699 = $log_{10}(50)$.

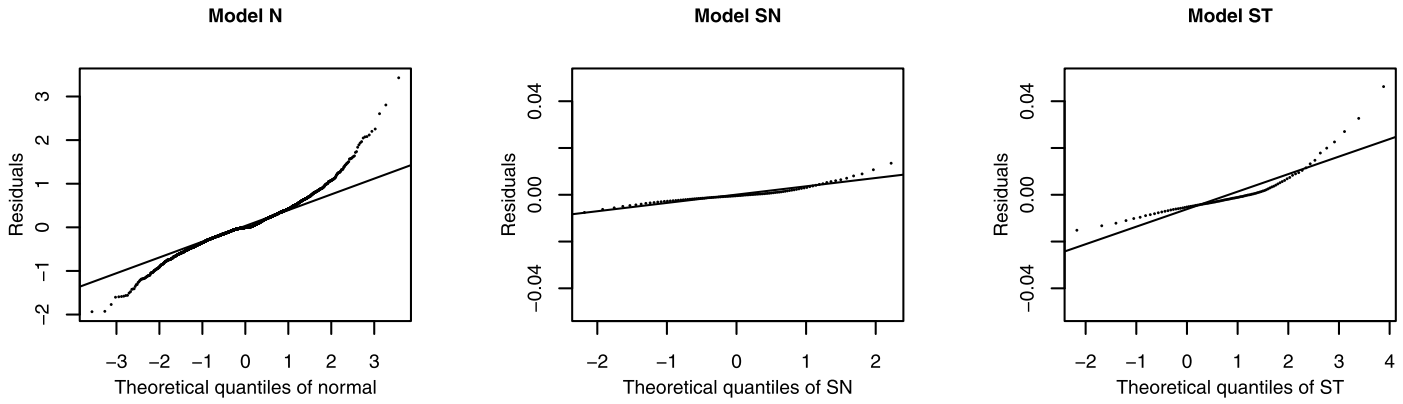

Figure 5. Normal, SN and ST Q–Q plots with line based on Models N, SN and ST.

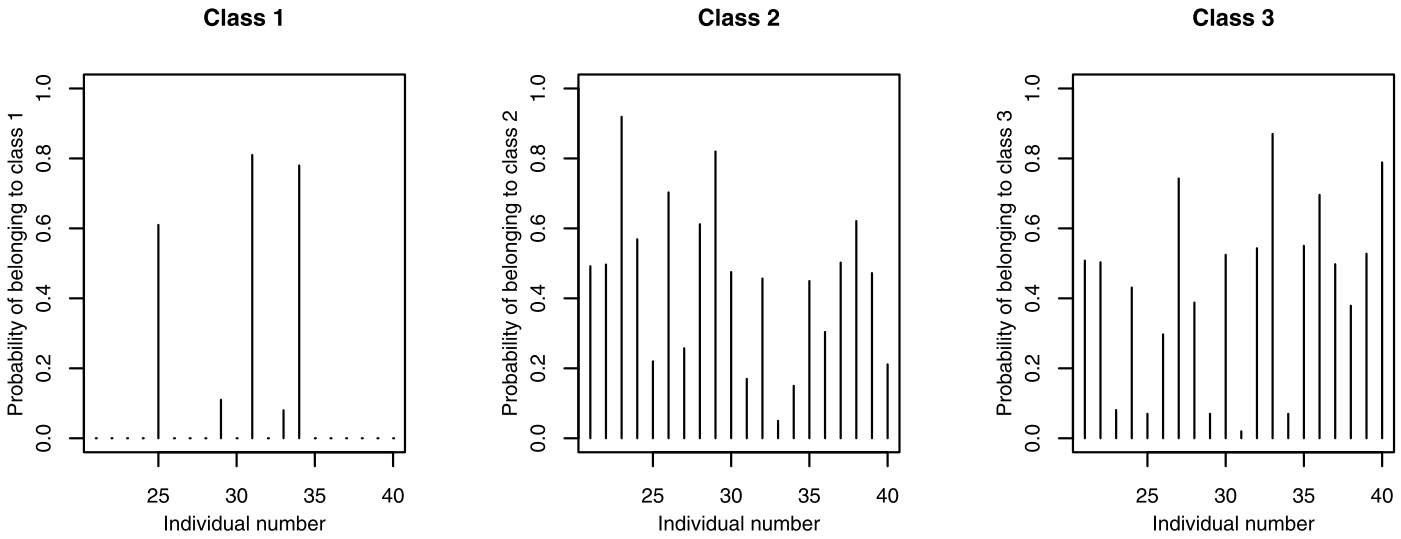

Figure 6. Posterior probabilities of belonging to 3 trajectory classes for the subjects 21 to 40 based on Model SN.

ship of individual i drawn from the posterior distribution (9) in the lth MCMC iteration ($l = 1, \ldots, M$), where M is total iteration number of posterior samples (=3,000 here). Barplot in Figure 6 displays the probabilities for the 20 individuals. The probability corresponding to individual patient who is classified as either viral load rebound or not may help physicians refine treatment strategy and to identify the potential reasons of viral load rebound for such individual patient. The individual posterior probabilities corresponding to the classes of trajectories shown in Figure 1(a) are matched quite well. That is, the patients 31 and 105 belong to class 1 because their viral load decrease constantly in a early short-term period, with probabilities 82% and 87%, respectively; the viral loads of the patients 29 and 132 decrease and then maintain stable, and thus, they belong to class 2, with probabilities 82% and 85%, respectively; and finally, the patients 33 and 99 are in class 3 (viral load rebound), with probabilities 87% and 100%.

The estimated population parameters presented in Table 1 based on Model SN indicate that the first-phase decay rate, and the second-phase decay rate with baseline CD4(0) and time-varying covariate CD4(t) may be approximated by $\hat{\lambda}_1 = 26.0 + 18.8\text{CD4}(0)$, $\hat{\lambda}_2 = -3.79$ and $\hat{\lambda}_2^*(t) = -3.79 + 12.5\widehat{\text{CD4}}(t)$, respectively, where $\widehat{\text{CD4}}(t)$ is the predicted time-varying CD4 covariate. Thus, the population viral load processes of 3 classes may be approximated by $\hat{V}_1(t) = \exp\left\{10.3 - \hat{\lambda}_1 t\right\}$, $\hat{V}_2(t) = \exp\left\{10.3 - \hat{\lambda}_1 t\right\} + \exp\left\{6.30 - \hat{\lambda}_2 t\right\}$ and $\hat{V}_3(t) = \exp\left\{10.3 - \hat{\lambda}_1 t\right\} + \exp\left\{6.30 - \hat{\lambda}_2^*(t) t\right\}$. Since the first-phase viral decay rate λ_1 is positively associated with the baseline CD4 (due to significant estimate of β_3) and the second-phase viral decay rate $\lambda_2^*(t)$ in component 3 is positively associated with the true (but unobserved) CD4 values, suggesting that the viral load $V(t)$ may be negatively significantly associated with both the baseline CD4 and the true CD4 values. This simple approximation considered here may provide a rough guidance and point to further research even though the true association described above may be more complicated.

In order to investigate how the measurement errors in CD4 contribute to modeling results, we further compare two methods for estimation based on Model SN: the proposed joint modeling approach and 'naive' method where the raw (observed) CD4 values, z_{ij}, rather than the true (unobservable) CD4 values, z_{ij}^*, are substituted in the mixture model. It can be seen from Tables 1 and 2 that there are important differences in the estimates for the parameters β_5 and β_6, which are directly associated with whether or not ignoring potential CD4 measurement errors for inference. The 'naive' method may produce unresonable estimates and substantially underestimate the covariate CD4 effect ('naive' method: $\hat{\beta}_6 = 2.27$ vs. joint modeling approach: $\hat{\beta}_6 = 12.5$). The estimated SD for the CD4 effect (β_6) using joint modeling approach is about three times larger than that using 'naive' method, probably because joint modeling approach incorporates the variation from fitting the CD4 process. The difference of estimates between joint modeling approach and 'naive' method, due to whether or not potential CD4 measurement errors are ignored for inference, indicates that CD4 measurement errors should not be ignored in the analysis. Thus, it is important to take the CD4 measurement errors into account when collected data are "imperfectly" measured.

We further investigated a commonly used NLMETJ model with SN distribution (Model NL, ignoring data feature of heterogeneous population), and compared it with the mixture of NLMETJ model (Model SN) to explore how heterogeneous feature influences modeling results. We found

Table 3. Summary of true parameter (TP) values, estimated $\boldsymbol{\beta}$, $\boldsymbol{\alpha}$ and skewness parameter δ as well as SD and MSE for Models N, SN and NL based on 150 generated data sets, under response model error with $\Gamma(2,1)$ distribution based on the joint modeling (JM) approach and the 'naive' method. Mean and SD denote averages of the MC posterior mean and MC standard deviation, respectively; MSE is quantified by percent $\sqrt{MSE} = 100 \times \sqrt{MSE_l}/|TP_l|$

	JM									NM		
	Model N			Model NL			Model SN			Model SN		
TP	Mean	SD	MSE	Mean	SD	MSE	Mean	SD	MSE	Mean	SD	MSE
$\beta_1 = 10.0$	10.82	0.27	6.54	10.48	0.27	4.24	10.12	0.21	2.65	10.33	0.24	3.81
$\beta_2 = 26.0$	28.54	4.45	8.23	27.13	4.89	7.46	26.12	2.76	3.17	26.66	4.98	5.96
$\beta_3 = 18.5$	20.14	5.49	10.23	19.34	6.11	9.12	18.87	2.77	5.81	19.03	3.98	8.06
$\beta_4 = 6.3$	7.31	2.53	9.34	7.01	1.42	6.39	6.41	0.89	2.23	6.85	1.20	5.80
$\beta_5 = -3.8$	-4.41	1.98	10.12	-4.12	1.77	9.43	-3.86	0.86	4.02	-4.34	2.23	10.11
$\beta_6 = 12.5$	9.33	6.77	10.65	10.82	6.47	9.23	12.91	3.45	5.66	9.29	7.33	11.02
$\alpha_1 = -0.2$	-0.17	0.43	4.33	-0.18	0.37	3.89	-0.19	0.35	3.67	-0.18	0.40	3.87
$\alpha_2 = 1.0$	1.10	2.21	3.89	1.09	1.63	2.98	1.03	0.77	1.02	1.08	1.32	1.87
$\alpha_3 = 6.5$	6.28	3.87	6.66	6.41	3.38	5.76	6.47	1.01	2.31	6.39	5.76	3.61
δ	–	–	–	1.34	0.84	–	1.52	0.97	–	1.46	0.96	–

the important difference in the estimates for the parameters (β_2, β_3), which are associated with the first-phase viral decay rate, and for the parameter (β_6), which is associated with the second-phase viral decay rate. We also obtained estimated DIC values with 11572.8 (Model SN) and 13031.1 (Model NL), respectively. Although Model SN has slightly larger EPD and SSR values than Model NL due to discrepancy of possible mis-classification, however, since Model SN estimated not only model parameters but also class membership probabilities, suggesting that the mixture modeling may be more efficient.

5. SIMULATION STUDIES

To assess the performance of the proposed mixture of NLMETJ models and methods, as an illustration, we conducted the following simulation studies where Models N, SN and NL are included and the joint modeling approach and 'naive' method are compared. To simulate a heterogeneous population, 40, 180 and 180 response trajectory mean values, out of 400 subjects, are generated based on the three functions in (19)–(21), respectively. We generated 150 datasets from specified models according to the additional specifications described below. The measurement time points used in the simulation are similar to those in the real data analysis and the true parameter values are mimic to those obtained in the example. $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_6)^T = (10.0, 26.0, 18.5, 6.3, -3.8, 12.5)^T$, $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \alpha_3)^T = (-0.2, 0.9, 6.5)^T$, $\boldsymbol{\Sigma}_b = \text{diag}(1.5, 6.6, 3.2, 9.9)$. The time-varying CD4 covariate z_{ij} are simulated from equation $z^*_{ij} = (-0.2 + a_{i1}) + (0.9 + a_{i2})t_{ij} + (6.5 + a_{i3})t^2_{ij}$, with $(a_{i1}, a_{i2}, a_{i3})^T \sim N(\mathbf{0}, \text{diag}(1.0, 0.5, 0.1))$ and ϵ_{ij} following N distribution $N(0, 0.02)$ for convenient implementation.

An advantage of the SN distribution for the model errors is its propensity for accommodating skewness. In the simulation, we generated e_{ij} according to $\Gamma(2,1)$ distribution, yielding a skewed distribution with $E(e_{ij}) = 2$ and $Var(e_{ij}) = 2$. Under this specification, data generated from mixture model with $\Gamma(2,1)$ distribution may exhibit highly skewed feature. We determined a threshold from generated data set as LOD value so that there are roughly 15% response observations below LOD. Note that the prior distributions considered are all close to non-informative as similarly treated in real data analysis. Thus, we expect the results to be somewhat robust with respect to prior distributions.

In the simulation studies, we investigated the following three scenarios. First, Model SN is compared with Model N to evaluate how skewness feature of data influences modeling results; second, Models SN and N are compared with Model NL to assess how data features with heterogeneous population and/or skewness simultaneously contribute to modeling results; and finally, for Model SN, we compared joint modeling approach with 'naive' method to investigate how the measurement errors in CD4 covariate affect to modeling results.

For evaluating the objective use of the criteria, the models preferred by DIC were recorded. For example, in the MCMC sampling result, none of DIC selected the N distribution specification for any of the 150 data sets, demonstrating the ability of selection method to detect an obvious departure from symmetry and suggesting strong evidence of skewness. Table 3 shows the numerical results for the estimates of parameters δ, $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$. The following findings were observed from the simulation study.

For the parameters in the covariate measurement error model, all parameters are similarly estimated in terms of models and methods, but α_1 and α_2 which are intercept and linear coefficient are overestimated, while α_3 which is quadratic coefficient is underestimated.

For the estimates of parameter vector $\boldsymbol{\beta}$, β_1 and β_4 were similar among the three fitted models and the two methods, while the estimates of the other parameters (β_2, β_3, β_5 and

β_6) tend to be significantly biased for Model N. We found that Models SN and NL work reasonably well in the current setting in terms of MSE (quantified by $100 \times \sqrt{\text{MSE}_l}/|\text{TP}_l|$, $l = 1, \ldots, 6$), while Model N may lead to large SDs and MSEs for most of parameters. Interestingly, for joint modeling approach we found that the SDs in Model N (symmetric) are larger than their counterparts in Models SN and NL (asymmetric), and the SDs in Model SN is smaller than corresponding those in Model NL, indicating that the model with longer tails and heterogeneous data feature may produce more precise estimates, and the efficiency of estimation on β_2, β_3, β_5 and β_6 associated with two phase decay rates is degraded (higher SD) when symmetric distribution for model error is assumed (Model N) or heterogeneous data feature is ignored (Model NL) relative to allowing a more flexible representation via both SN distributions and heterogeneous data feature (Model SN). This suggests that adopting the (symmetric) normal distribution assumption or ignoring heterogeneous data feature may lead to inaccurate and inefficient inference on fixed-effects of primary interest, in particular, when data exhibit asymmetric and/or heterogeneous features. Thus, for Models N, NL and SN, Model SN performs the best, followed by Model NL and Model N is worst. In addition, MC means of the skewness parameters of Models NL and SN are 1.34 and 1.52, respectively, indicating the data departs from N distribution.

We can also see from the simulation results in Table 3, that 'naive' method method may lead to biased estimates and large MC SD and MSEs for all parameters in comparison with joint modeling approach, in particular, for β_5 and β_6 which are closely related to whether measurement errors in CD4 cell count are ignored or not. Therefore, it is important to model the data incorporating measurement errors in CD4 covariate into account.

6. DISCUSSION

For longitudinal biomarkers, viral load and CD4 count used as outcome measures to evaluate treatment effects of ART in AIDS clinical trials, to understand pathogenesis of HIV infection and to assess risk of disease progression to AIDS, we have developed a simultaneous Bayesian modeling approach to the finite mixture of NLMETJ models for longitudinal viral load data with features of heterogeneous population, non-normality and left-censoring due to LOD in the presence of mismeasured CD4 covariate for estimating both model parameters and class membership probabilities. Along with this line, one advantage of mixture modeling is its flexibility to handle longitudinal data with different characteristics and provides not only estimates of all model parameters, but also model-based probabilistic clustering to obtain class membership probabilities at both population and individual levels. This information may help clinicians refine general treatment strategy and develop individualized ART regimens. This kind of mixture modeling approach is important in many biostatistical application areas, allowing accurate inference of parameters while adjusting for heterogeneity, non-normality and incompleteness in the longitudinal data. Although this paper is motivated by AIDS clinical study, the basic concepts of the developed mixture of NLMETJ models have generally broader applications whenever the relevant technical specifications are met and longitudinal measurements are assumed to arise from two or more identifiable subclasses within a population.

This article have considered two mixture modeling methods to compare potential models with different specification of error distributions. In particular, (i) we investigated how asymmetric error distributions (Models SN and ST) contribute to inferential results in comparison with a symmetric (normal) error distribution (Model N). Our results support that it is important to assume a skew distribution in the mixture joint models in order to achieve reliable results, in particular if the data exhibit the feature of non-normality. (ii) We compared joint modeling approach with 'naive method' to investigate how the measurement errors in CD4 affect modeling results and inference. The findings indicate that it is critical to take the CD4 measurement errors into account when collected data are "imperfectly" measured. (iii) Under the assumption of SN distribution, we further explored how heterogeneous data feature influences modeling results by comparing Model SN with a commonly used NLMETJ model (Model NL), in which the data feature of heterogeneous population is ignored. The results suggest that there are important differences due to whether the data feature of heterogeneous population is ignored or not for inference. The proposed mixture joint models and methods may have a significant impact on HIV/AIDS research and ART to AIDS patients, and help improve understanding of the pathogenesis of HIV infection and evaluation of individualized ART regimens.

For inference of mixture models, parameter (or model) identifiability can be an important but difficult problem when a large number of model parameters are estimated simultaneously. We must ensure each component model to be identifiable to make whole mixture model to be identifiable. To make (20) and (21) to be identifiable, we assume $\lambda_{1i} > \lambda_{2i}$ in (20) and $\lambda_{1i} > \lambda^*_{2ij}$ in (21). In practice, if the models are not identifiable, the MCMC algorithm would diverge quickly. In the application considered in this article, the MCMC algorithm was converged without problems and we did not observe potential identifiability problems.

It is noted that a fundamental problem for "traditional" Bayesian mixture model analysis, in which each component has the same family of densities but with different sets of parameters, is label switching due to the non-identifiability of mixture components [27]. However, fortunately, this problem did not happen in this case because components correspond to different densities with distinguished mean functions, g_k $(j = 1, 2, 3)$ in our case, which are known and prespecified. As noted by [23], to formularize the finite mixture models, the mean functions of component submodels can be similar in form, but varying only mean and/or variance specifications or have entirely different functional forms with

parameters of different dimensions and meanings across the component submodels, which is the case conducted in this article; see equations (19)–(21). Instead of specifying different mean functions for classes, alternatively, one can also specify an universal mean function for all classes, for example, mean function (21), and let data themselves determine the number of clusters and/or shapes of trajectories. In that way the label switching issue may arise and some labelling methods [27, 34] must be applied to solve this problem, but significantly additional efforts are needed based on the finite mixture of joint models proposed in our study. We are actively investigating these important issues, and hope that these interesting results could be reported in the near future.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge one anonymous referee, an Associate Editor and an Editor for their insightful comments and constructive suggestions that led to a marked improvement of the article. Y. Huang was partially supported by University South Florida Proposal Enhancement grant (18326); J. Chen was partially supported by the National Natural Science Foundation of China (81671633) and the Natural Science Founds of Hubei Province (2014CFB863); and P. Yin was partially supported by the National Natural Science Foundation of China (81573262).

Received 22 May 2016

REFERENCES

[1] Arellano-Valle, R. B. and Genton, M., On fundamental skew distributions. *Journal of Multivariate Analysis* 2005; **96**: 93–116. MR2202402

[2] Arellano-Valle, R. B., Bolfarine, H., and Lachos, V. H., Bayesian inference for skew-normal linear mixed models. *Journal of Applied Statistics* 2007; **34**: 663–682. MR2410041

[3] Azzalini, A. and Genton, M., Robust likelihood methods based on the skew-t and related distributions. *International Statistical Review* 2008; **76**: 106–129.

[4] 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. MR2243417

[5] Dagne, G. A. and Huang, Y., Bayesian semiparametric mixture Tobit models with left censoring, skewness, and covariate measurement errors. *Statistics in Medicine* 2013; **32**: 3881–3898. MR3102446

[6] Davidian, M. and Giltinan, D. M., 1995. *Nonlinear Models for Repeated Measurement Data*. London, Chapman and Hall.

[7] Diebolt, J. and Robert, C., Estimation of finite mixture distributions by Bayesian sampling. *J. Royal Statist. Soc. Series B* 1994; **56**: 363–375. MR1281940

[8] Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B., 2003. *Bayesian Data Analysis*. London: Chapman and Hall. MR2027492

[9] Gelman, A. and Rubin, D., Inference from iterative simulation using multiple sequences. *Statistical Science* 1992; **7**: 457–511.

[10] Hammer, S. M., Vaida, F., Bennett, K. K., Holohan, M. K., Sheiner, L., Eron, J. J., Wheat, L. J., Mitsuyasu, R. T., Gulick, R. M., Valentine, F. T., Aberg, J. A., Rogers, M. D., Karol, C. N., Saah, A. J., Lewis, R. H., Bessen, L. J., Brosgart, C., DeGruttola, V., and Mellors, J. W., Dual vs single protease inhibitor therapy following antiretroviral treatment failure: a randomized trial. *The Journal of the American Medical Association* 2002; **288**: 169–180.

[11] Higgins, K. M., Davidian, M., and Giltinan, D. M., A two-step method to measurement error in time-dependent covariates in nonlinear mixedeffects models, with application to IGF-I pharmacokinetics. *Journal of the American Statistical Association* 1997; **92**: 436–448.

[12] Ho, H. J. and Lin, T. I., Robust linear mixed models using the skew-t distribution with application to schizophrenia data. *Biometrical Journal* 2010; **52**: 449–469. MR2751913

[13] Huang, Y., Liu, D., and Wu, H., Hierarchical Bayesian methods for estimation of parameters in a longitudinal HIV dynamic system. *Biometrics* 2006; **62**: 413–423. MR2227489

[14] Huang, Y., Chen, J., and Yan, C., Mixed-effects joint models with skew-normal distribution for HIV dynamic response with missing and mismeasured time-varying covariate. *International Journal of Biostatistics* 2012; **8**: 1–30. MR3003819

[15] Huang, Y. and 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. MR2898838

[16] Jara, A., Quintana, F., and Martin, E. S., Linear mixed models with skew-elliptical distributions: a Bayesian approach. *Computational Statistics and Data Analysis* 2008; **52**: 5033–5045. MR2526212

[17] Lavine, M. and West, M., A Bayesian method for classification and discrimination. *Canadian Journal of Statisttics* 1992; **20**: 451–461. MR1208356

[18] Liu, W. and Wu, L., Simultaneous inference for semiparametric nonlinear mixed-effects models with covariate measurement errors and missing responses. *Biometrics* 2007; **63**: 342–350. MR2370792

[19] Lu, X. and Huang, Y., Bayesian analysis of nonlinear mixed-effects mixture models for longitudinal data with heterogeneity and skewness. *Statistics in Medicine* 2014; **33**: 2830–2849. MR3256541

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

[21] Muthén, B. and Shedden, K., Finite mixture modeling with mixture outcomes using the EM algorithm. *Biometrics* 1999; **55**: 463–469.

[22] Nowak, M. A. and May, R. M., 2000. *Virus Dynamics: Mathematical Principles of Immunology and Virology*. Oxford: Oxford University Press. MR2009143

[23] Pauler, D. and Laird, N. M., A mixture model for longitudinal data with application to assessment of noncompliance. *Biometrics* 2000; **56**: 464–472.

[24] Perelson, A. S., Essunger, P., Cao, Y., Vesanen, M., Hurley, A., Saksela, K., Markowitz, M., and Ho, D. D., Decay characteristics of HIV-1-infected compartments during combination therapy. *Nature* 1997; **387**: 188–191.

[25] Sahu, S. K., Dey, D. K., and Branco, M. D., A new class of multivariate skew distributions with applications to Bayesian regression models. *The Canadian Journal of Statistics* 2003; **31**: 129–150. MR2016224

[26] Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and Van der Linde, A., Bayesian measures of model complexity and fit. *Journal of the Royal Statiscal Socity, Series B* 2002; **64**: 583–639. MR1979380

[27] Stephens, M., Dealing with label switching in mixture models. *Journal of the Royal Statiscal Socity, Series B* 2000; **62**: 795–809. MR1796293

[28] Titterington, D. M., Smith, A. F. M., and Makov, U. E., 1985. *Statistical Analysis of Finite Mixture Distributions*. Chichester, U.K.: John Wiley and Sons. MR0838090

[29] Tobin, J., Estimation of relationships for limited dependent variables. *Econometrica* 1958; **26**: 24–36. MR0090462

[30] Wu, H. and Ding, A. A., Population HIV-1 dynamics in vivo: applicable models and inferential tools for virological data from AIDS clinical trials. *Biometrics* 1999; **55**: 410–418.
[31] Wu, L., A joint model for nonlinear mixed-effects models with censoring and covariates measured with error, with application to AIDS studies. *Journal of the American Statistical Association* 2002; **97**: 955–964. MR1951254
[32] Wu, H. and Wu, L., Identification of significant host factors for HIV dynamics modeled by nonlinear mixed-effects models. *Statistics in Medicine* 2002; **21**: 753–771.
[33] Wu, H. and Zhang, J. T., 2006. *Nonparametric Regression Methods for Longitudinal Data Analysis*. Hoboken, New Jersey: Wiley. MR2216899
[34] Yao, W. and Lindsay, B. G., Bayesian mixture labeling by highest posterior density. *Journal of American Statistical Association* 2009; **104**: 758–767. MR2751453

Yangxin Huang
Department of Epidemiology and Biostatistics
College of Public Health
University of South Florida
Tampa, Florida 33612
USA

Department of Statistics
College of Science
Wuhan University of Technology
Wuhan, Hubei 430070
People's Republic of China

Department of Mathematics
School of Mathematics and Computer
Wuhan Textile University
Wuhan, 430073
People's Republic of China
E-mail address: yhuang@health.usf.edu

Jiaqing Chen
Department of Statistics
College of Science
Wuhan University of Technology
Wuhan, Hubei 430070
People's Republic of China
E-mail address: jqchenwhut@163.com

Ping Yin
Department of Epidemiology and Biostatistics
School of Public Health
Huazhong University of Science and Technology
Wuhan, Hubei 430030
People's Republic of China
E-mail address: pingy2000@163.com

Huahai Qiu
Department of Mathematics
School of Mathematics and Computer
Wuhan Textile University
Wuhan, 430073
People's Republic of China
E-mail address: qiuhuahai2006@163.com