Skew-Elliptical Mixed-Effects Models with Time-Dependent Viral Decay Rates

Yangxin Huang* and Ren Chen *

Abstract

Mixed-effects models with different time-varying decay rate functions have been proposed for HIV/AIDS studies. However (i) it is not clear which model is more appropriate, and (ii) the model random error is commonly assumed to follow a normal distribution, which may be unrealistic and can obscure important features of within- and among-subject variations. Because asymmetry of HIV viral load data is still noticeable even after transformation, it is important to use a more general distribution family that enables the unrealistic normal assumption to be relaxed. We developed skewed elliptical (SE) Bayesian mixed-effects models by considering the model random error to have an SE distribution. We compared five SE models that have different time varying decay rate functions. For each model, we also contrasted the performance under different model random error assumption such as normal, Student-t, skew-normal or skew-t distribution. The results indicate that the model with a time-varying viral decay rate that has two exponential components is preferred. The models with skew distributions have been shown beneficial in dealing with asymmetric data and provided better fitting to the data than those with symmetric distributions.

Key Words: Bayesian analysis, HIV/AIDS, mixed-effects models, skew-elliptical distribution, time-dependent viral decay rate

1. Introduction

Mathematical modeling is an important tool for understanding the evolution of HIV viral load (number HIV-1 RNA copies in plasma) and interactions between HIV and its target cells. The viral load trajectory is complex and has multiple phases of change (Maldarelli et al., 2007; Perelson et al., 1997). Data from the AIDS clinical trial, ACTG5055 (Acosta et al., 2004) in Figure 1(a), show that: (i) within the first 2 weeks after the initial treatment, the viral load (transformed in natural log scale) dropped linearly and sharply, therefore, the change of viral load can be approximated by an exponential function; (ii) within the first 2–3 months but after the first 2 weeks, the relationship between the viral load and time was still linear but the slope became flatter, which indicates a slower decay rate; (iii) between the third to eighth month, the viral load either decreased more slowly, remained at a constant low level, or started to increase up to the level measured before treatment was initiated. The possible reasons for viral load rebound are development of resistance to the medications, and other clinical issues such as lack of adherence. There is no clear cutoff among the phases, not every subject will have all of these phases and the length of the phases may vary among individuals. Therefore, the associated decay rate in the models for the viral load trajectories is expected to vary over time and can be individually specific.

For the first phase of HIV viral load dynamics (i.e., the first 1 to 2 weeks), we can apply a uniexponential equation (Ho et al., 1995; Wei et al., 1995) as,

$$V(t) = V(0)\exp(-\lambda t), \tag{1}$$

where $V(t)$ is total viral load at time t, $V(0)$ is the baseline viral load at $t = 0$ and λ is a constant viral change rate which is the speed of the loss of viral load after initiation of antiviral treatment. Although equation (1) can precisely describe the phenomenon of a

*Department of Epidemiology & Biostatistics, College of Public Health, MDC 56, University of South Florida, Tampa, FL 33612, USA

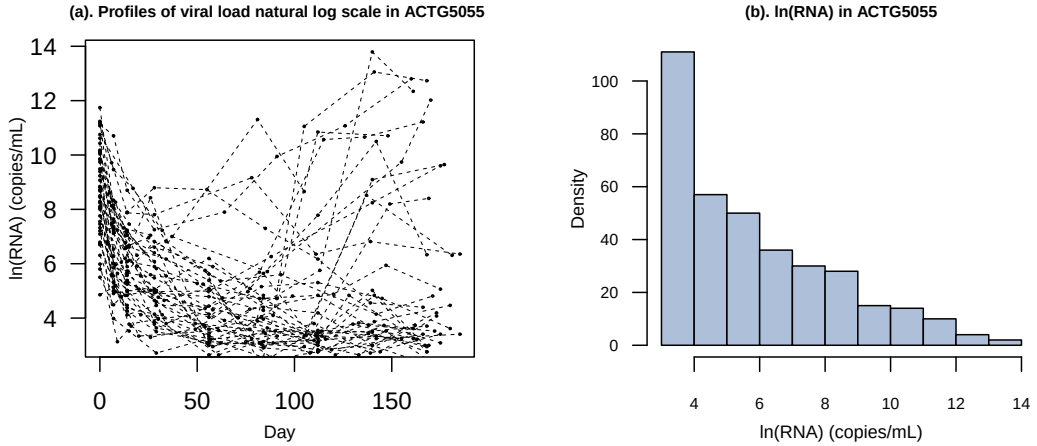

Figure 1: Profile and histogram of viral load in natural log scale from a clinical trial study.

linear decrease of logarithm transformed viral load within approximately one to two weeks since treatment, we cannot apply it to the whole trajectory because the viral load is only allowed to decrease at a constant rate in this equation. Besides that, there are at least three unsolved issues.

First, in order to use entire HIV follow up data, extended from equation (1), different models have been proposed in the literature, it is unclear which one is more appropriate.

Second, in mixed-effects model for longitudinal data analysis, random errors and/or random-effects are usually assumed to have a normal distribution. Although the normality assumption is satisfied in many situations, it may cause biased or misleading inference if the data include extreme values or show skewness with heavy tails, which are commonly seen in virological responses (Huang and Dagne, 2011; Sahu et al., 2003; Verbeke and Lesaffre, 1996). Figure 1(b) displays the histogram of repeated viral load in natural log scale for 44 subjects enrolled in the ACTG5055 study. The skew pattern, which is still obvious even after the transformation, tends to be right and ranges from 0.8 and 2.15 at each of the follow up measurements. If the ratio between skewness value and standard error of skewness is greater than 2, the data may be regarded as having unignorable skewness (Gardner, 2001). In the ACTG5055 study, the ratio with 4 indicates that skewness needs to be accounted for.

Third, computation can be a challenge. Frequentist and Bayesian are two major approaches used in studies of HIV dynamics. In the frequentist approach, based on the maximum likelihood estimation (MLE), different extensions have been proposed, such as Laplace approximation of the numerical integrals (Beal and Sheiner, 1982; Lindstrom and Bates, 1990; Wu and Zhang, 2002), stochastic approximation EM (SAEM) algorithm (Kuhn and Lavielle, 2005; Lavielle et al., 2011), joint model via Monte Carlo EM algorithm (Liu and Wu, 2007; Wu, 2004) and asymptotic distribution of the maximum h-likelihood estimators (MHLE) (Commenges et al., 2011). The second approach is Bayesian mixed-effects modeling via Markov chain Monte Carlo (MCMC) (Huang et al., 2006; Huang and Dagne, 2011; Putter et al., 2002). The Bayesian approach is an efficient way to incorporate prior information, both point estimates and uncertainties (variances), into analysis to identify more unknown parameters in complex models.

The main focus of this paper is to provide a comprehensive comparison of five commonly used HIV dynamic models with skew-elliptical (SE) distribution (Sahu et al., 2003) in random error under Bayesian approach. The rest of the paper is organized as follows. Section 2 presents the HIV dynamic models that have a time-varying decay rate function, so they can be applied to the entire follow-up data. In Section 3, we explore a general Bayesian mixed-effects modeling approach. In Section 4, we describe the motivated AIDS

data and present results of model comparisons. Finally, the paper concludes with some discussions in Section 5.

2. HIV Dynamic Models with Time-Varying Viral Decay Rate Function

As mentioned in Section 1, there is a multiphasic change in HIV viral load after the initiation of HAART. One potential interpretation of this phenomenon is that the process involves distinct populations with different homogenous behaviors. For example, the fast decreasing decay rate observed in the first phase is due to the treatment effect on productively infected CD4 cells, while the slower decay rate in the second phase is primarily due to the effect on the latently or long-lived infected cells (Perelson et al., 1997). However, some phenomena can't be explained by this theory. For example, there can be large differences in mean decay rates in response to different treatment regimens: during the first week, the death of infected cells may be substantially slower during days 3–6 than during days 2–3 (Grossman et al., 1999). Following equation (1), Grossman et al.(1999) proposed an equation for viral load as:

$$\begin{aligned} V(t) &= V(0)\exp\{-\tfrac{\log(R)}{\tau}t\}, \\ R(v) &= R(1)+\alpha(1-v)\exp(-\rho v), \\ v(t) &= V(t)/V(0), \end{aligned} \tag{2}$$

where R is reproduction ratio and is a function of time, ρ is an adjustable parameter, τ is the average infection cycle time, $v(t)$ is the ratio between the viral load at time *t* and the baseline. Following Zhang and Wu (2011), equation (2) is equivalent to the following equation,

$$V(t) = V(0)\exp(-\lambda(t)t), \tag{3}$$

where $\lambda(t) = \int_0^t \alpha(\tau)d\tau/t$ is the average relative lost rate of the viral load $V(t)$. $\lambda(t)$ can be positive (if $R < 1$), zero (if $R = 1$), or negative (if $R > 1$). If $\lambda(t) < 0$, the decay rate $\lambda(t)$ at time *t* actually is a growth rate. Therefore, by including a time-varying decay rate function $\lambda(t)$, compared with equation (1), equation (3) is more flexible and can be applied to include the entire follow up data without the need to arbitrarily truncate the data.

A unified model with a time-varying viral decay rate function can be expressed as:

$$y(t) = \ln\{V(0)\exp[-\lambda(t)t]\} + \epsilon = \beta_1 - \lambda(t)t + \epsilon, \tag{4}$$

where $y(t)$ is the natural logarithm transformation of the number of HIV-1 RNA copies per mL of plasma, ϵ is the measurement error, $\ln(V(0)) = \beta_1$ and is the macro-parameter for initial viral load in natural $\log$ scale.

Among the different decay rate functions proposed in literature, we select five representatives as follow (Dagne and Huang, 2012; Grossman et al., 1999; Wu, 2004; Zhang and Wu, 2011),

$$\begin{aligned} I: &\ \lambda(t) = \beta_2 + \beta_3 t, \\ II: &\ \lambda(t) = \beta_2\exp(-\beta_3 t) + \beta_4, \\ III: &\ \lambda(t) = \beta_2\exp(-\beta_3 t) + \beta_4 + \beta_5 t, \\ IV: &\ \lambda(t) = \beta_2\exp(-\beta_3 t) + \beta_4\exp(-\beta_5 t), \\ V: &\ \lambda(t) = v[w(t), h_i(t)], \end{aligned} \tag{5}$$

where the last one, $\lambda(t) = v[w(t), h_i(t)]$, is a nonparametric function.

3. Bayesian Mixed-Effects Models with Skewed Distributions

To account for the skewness observed in the data, the random errors in mixed-effects models can be assumed to follow an SE distribution (Huang and Dagne, 2010, 2011; Sahu etal, 2003). The SE distribution is a family of distributions that is not only mathematically

tractable but also flexible in its possible shapes. Because in the SE family, skew-normal (SN), normal and Student-*t* distribution are all a special case of skew-*t* (ST) distribution, therefore, in this section, we present a general form of a mixed-effects model with an ST distribution under the Bayesian approach. A general mixed-effects model with an ST distribution can be expressed as,

$$\begin{array}{rcll} \boldsymbol{y}_i & = & \boldsymbol{g}_i(\boldsymbol{t}_i, \boldsymbol{\beta}_i) + \boldsymbol{\epsilon}_i, & \boldsymbol{\epsilon}_i \overset{\text{iid}}{\sim} ST_{n_i,\nu}(0, \boldsymbol{\Sigma}, \boldsymbol{\Delta}), \\ \boldsymbol{\beta}_i & = & \boldsymbol{d}(\boldsymbol{\beta}, \boldsymbol{b}_i), & \boldsymbol{b}_i \overset{\text{iid}}{\sim} N(\boldsymbol{0}, \boldsymbol{\Sigma}_b), \end{array} \tag{6}$$

$\boldsymbol{y}_i = (y_{i1}, \ldots, y_{in_i})^T$ with y_{ij} being the response value for the ith individual at the jth time $(i = 1, ..., n; j = 1, ..., n_i)$, $\boldsymbol{g}_i(\boldsymbol{t}_i, \boldsymbol{\beta}_i) = (g(t_{i1}, \boldsymbol{\beta}_i), \ldots, g(t_{in_i}, \boldsymbol{\beta}_i))^T$, $\boldsymbol{t}_i = (t_{i1}, \ldots, t_{in_i})^T$, $\boldsymbol{\beta}_i$ and $\boldsymbol{\beta}$ are individual-specific parameter vector and population parameter vector, respectively, $\boldsymbol{g}(\cdot)$ and $\boldsymbol{d}(\cdot)$ are linear or nonlinear known parametric functions, $\boldsymbol{b}_i$ is normal random-effect vector with $\boldsymbol{\Sigma}_b$ being an unstructured covariance matrix. The vector of random errors $\boldsymbol{\epsilon}_i = (\epsilon_{i1}, \ldots, \epsilon_{in_i})^T$ follows a multivariate ST distribution with degrees of freedom ν, within-subject covariance matrix $\boldsymbol{\Sigma}$ and we usually can assume $\boldsymbol{\Sigma} = \sigma^2 \boldsymbol{I}_{n_i}$, and unknown $n_i \times n_i$ skewness diagonal matrix such that $\boldsymbol{\Delta} = diag(\delta_{i1}, \ldots, \delta_{in_i})$, skewness parameter vector $\boldsymbol{\delta}_i = (\delta_{i1}, \ldots, \delta_{in_i})^T$. In particular, if $\delta_{i1} = \cdots = \delta_{in_i} \hat{=} \delta$, then $\boldsymbol{\Delta} = \delta \boldsymbol{I}_{n_i}$ and $\boldsymbol{\delta}_i = \delta \boldsymbol{1}_{n_i}$, where $\boldsymbol{1}_{n_i} = (1, \ldots, 1)^T$, indicating that we are interested in skewness of overall data set.

Following discussion by Huang and Dagne (2011) and Sahu et al.(2003), to implement an MCMC procedure to model (6), by introducing one $n_i \times 1$ random vector $\boldsymbol{w}_i$, based on the stochastic representation, the model can be hierarchically formulated as follows.

$$\begin{array}{rcl} \boldsymbol{y}_i | \boldsymbol{b}_i, \boldsymbol{w}_i & \sim & t_{n_i,n_i+\nu}(\boldsymbol{g}_i(\boldsymbol{t}_i, \boldsymbol{\beta}_i) + \delta \boldsymbol{w}_i, w_i \sigma^2 \boldsymbol{I}_{n_i}), \\ \boldsymbol{w}_i & \sim & t_{n_i,n_i+\nu}(\boldsymbol{0}, \boldsymbol{I}_{n_i}) I(\boldsymbol{w}_i > \boldsymbol{0}), \\ \boldsymbol{b}_i & \sim & N(\boldsymbol{0}, \boldsymbol{\Sigma}_b), \end{array} \tag{7}$$

where $w_i = (\nu + \boldsymbol{w}_i^T \boldsymbol{w}_i)/(\nu + n_i)$, $t_{n_i,\nu}(\boldsymbol{\mu}, \boldsymbol{A})$ denotes the n_i-variate Student-*t* distribution with parameters $\boldsymbol{\mu}$, $\boldsymbol{A}$ and degrees of freedom ν, $I(\boldsymbol{w} > \boldsymbol{0})$ is an indicator function and $\boldsymbol{w} = |\boldsymbol{X}_0|$ with $\boldsymbol{X}_0 \sim t_{n_i,\nu}(\boldsymbol{0}, \boldsymbol{I}_{n_i})$. Note that the hierarchical models above under Bayesian framework will allow researchers to easily implement the methods using the freely available WinBUGS software (Lynn et al., 2000) and the computational effort for the model with an ST distribution is almost equivalent to that of the model with a Student-*t* distribution.

The unknown population parameters in the model (6) are $\boldsymbol{\theta} = \{\boldsymbol{\beta}, \sigma^2, \boldsymbol{\Sigma}_b, \nu, \delta\}$, and we assume they are independent of one another. Under Bayesian framework, we also need to specify prior distributions for unknown parameters as follows.

$$\boldsymbol{\beta} \sim N(\beta_0, \boldsymbol{\Lambda}), \sigma^2 \sim IG(\omega_1, \omega_2), \boldsymbol{\Sigma}_b \sim IW(\boldsymbol{\Omega}, v), \nu \sim Exp(\nu_0) I(\nu > 2), \delta \sim N(\boldsymbol{0}, \gamma), \tag{8}$$

where the mutually independent Normal (N), Inverse Gamma (IG), Exponential (Exp) and Inverse Wishart (IW) prior distributions are chosen to facilitate computations (Davidian and Giltinan, 1995). The super-parameter matrices $\boldsymbol{\Lambda}$ and $\boldsymbol{\Omega}$ can be assumed to be diagonal for convenient implementation.

Let $\pi(.)$ be a prior density function, so $\pi(\boldsymbol{\theta}) = \pi(\boldsymbol{\beta})\pi(\sigma^2)\pi(\boldsymbol{\Sigma}_b)\pi(\nu)\pi(\delta)$. Denote the observed data by $\mathcal{D} = \{\boldsymbol{y}_i, i = 1, ..., n\}$, and $f(\cdot|\cdot)$ as a conditional density function. Based on Bayesian inference, the posterior density of $\boldsymbol{\theta}$ is proportional to the observed data and prior distribution as:

$$f(\boldsymbol{\theta}|\mathcal{D}) \propto \{\prod_i^n \int f(\boldsymbol{y}_i | \boldsymbol{b}_i, \boldsymbol{w}_i; \boldsymbol{\beta}, \sigma^2, \nu, \delta)\ f(\boldsymbol{w}_i | \boldsymbol{w}_i > \boldsymbol{0})\ f(\boldsymbol{b}_i | \boldsymbol{\Sigma}_b) d\boldsymbol{b}_i\} \pi(\boldsymbol{\theta}). \tag{9}$$

In general, the integral in (9) is of high dimension and does not have any closed form. Analytic approximations to the integral 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) by the Gibbs sampling along with the Metropolis-Hastings (M-H) algorithm.

4. AIDS Clinical Data Analysis

4.1 Data Description and Specific Models

We used two AIDS clinical trials to explore the best fit among the models with different time-varying decay rate functions and different model error assumption such as normal, SN, Student-*t* and ST distributions. The first trial, ACTG5055 (Acosta et al., 2004), is the focus. Further, we used data from another clinical trial, ACTG398 (Pfister et al., 2003), to validate the conclusions obtained from ACTG5055.

ACTG5055 study was a phase I/II, randomized, open-label, 24-week comparative study. It included 44 HIV-1 infected patients who failed their first protease inhibitor (PI) treatment. Subjects were randomly assigned into one of the two arms. RNA viral load was measured (copies$/mL$) in blood samples collected at study days 0, 7, 14, 28, 56, 84, 112, 140 and 168 with a low limit of quantification of 50 copies$/mL$. HIV-1 RNA measures below this limit are not considered reliable. Following the simple substitution method of one half the detection limit for the detection (Helsel, 1990), we used 25 copies$/mL$ if the viral load was 50 copies$/mL$ or less.

ACTG398 study was a phase II trial that included 481 HIV-1 positive patients with prior exposure to approved PIs and loss of virological suppression. All patients were assigned to receive routine antiretroviral treatment (ART). Besides these medications, depending on the dose and type of PIs to which the patients previously exposed, they were selectively randomly assigned into one of four groups. HIV-1 RNA levels were measured at the time of entry into the study (day 0), at study weeks 2, 4, 8, 16, 24, 32, 40, and 48, every 8 weeks thereafter, and at the time of confirmed virological failure. The low limit of quantification is 100 copies$/mL$ and the HIV-1 RNA measures below this limit are not considered reliable and 50 copies$/mL$ was used instead. We draw two samples from ACTG398 based on the method of simple random sampling without replacement, one sample includes 44 subjects and the other includes 100 subjects. We also used all of the 481 subjects in ACTG398 in the model comparisons.

Figure 2 shows the measurements of viral load in natural $\log$ scale for four randomly selected patients from ACTG5055 and two sample data sets from ACTG398. We can see that viral load trajectories vary widely and they are substantially different across individuals. To account for this time-varying viral load change, we applied a mixed-effects model with a time-varying decay rate function, as discussed in Section 2. In addition, we assumed the model errors followed an ST distribution in order to make it flexible in dealing with the skewness observed in the data. The exact day of viral load measurement was used to compute study day in our analysis.

Under the general layout as model (4), corresponding to the five time-varying decay functions presented in Section 2, the mixed-effects models can be expressed as follow.

Model I: Quadratic linear mixed-effects model:

$$\begin{array}{rcl} y_{ij} & = & \beta_{1i} - [\beta_{2i} + \beta_{3i} t_{ij}] t_{ij} + e_{ij}, \\ \boldsymbol{e}_i & \overset{\text{iid}}{\sim} & ST_{n_i,v}(0, \sigma^2 \boldsymbol{I}_{n_i}, \delta \boldsymbol{I}_{n_i}), \\ \beta_{1i} & = & \beta_1 + b_{1i}, \;\; \beta_{2i} = \beta_2 + b_{2i}, \;\; \beta_{3i} = \beta_3 + b_{3i}, \end{array} \tag{10}$$

where $\boldsymbol{\beta} = (\beta_1, \beta_2, \beta_3)^T$ and $\boldsymbol{b}_i = (b_{1i}, b_{2i}, b_{3i})^T \overset{\text{iid}}{\sim} N_3(\mathbf{0}, \boldsymbol{\Sigma}_{\mathbf{b}})$

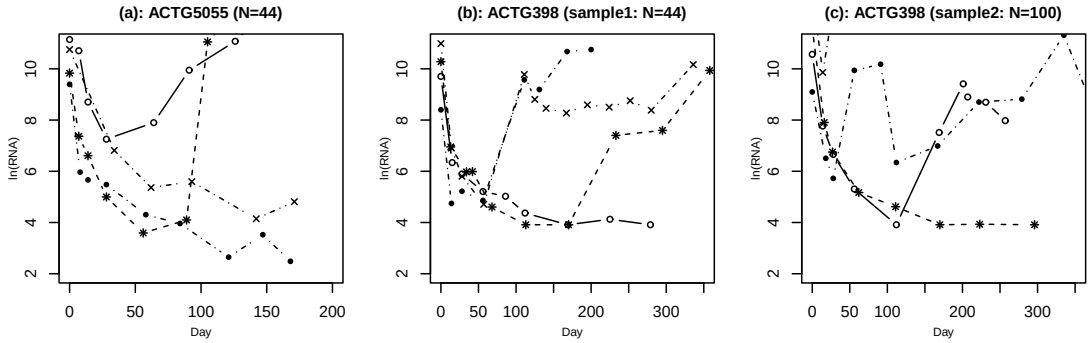

Figure 2: Profiles of viral load in natural log scale for four randomly selected patients among ACTG5055 and ACTG398, respectively.

Model II: Nonlinear mixed-effects model (uniexponential plus a constant):

$$\begin{array}{rcl} y_{ij} & = & \beta_{1i} - [\beta_{2i}\exp(-\beta_{3i}t_{ij}) + \beta_{4i}]t_{ij} + e_{ij}, \\ \boldsymbol{e}_i & \overset{\text{iid}}{\sim} & ST_{n_i,v}(0, \sigma^2\boldsymbol{I}_{n_i}, \delta\boldsymbol{I}_{n_i}), \\ \beta_{1i} & = & \beta_1 + b_{1i},\ \ \beta_{2i} = \beta_2 + b_{2i},\ \ \beta_{3i} = \beta_3 + b_{3i},\ \ \beta_{4i} = \beta_4 + b_{4i}, \end{array} \tag{11}$$

where $\boldsymbol{\beta} = (\beta_1, \beta_2, \beta_3, \beta_4)^T$ and $\boldsymbol{b}_i = (b_{1i}, b_{2i}, b_{3i}, b_{4i})^T \overset{\text{iid}}{\sim} N_4(\mathbf{0}, \boldsymbol{\Sigma}_\mathbf{b})$

Model III: Nonlinear mixed-effects model (uniexponential plus a linear function):

$$\begin{array}{rcl} y_{ij} & = & \beta_{1i} - [\beta_{2i}\exp(-\beta_{3i}t_{ij}) + \beta_{4i} + \beta_{5i}t_{ij}]t_{ij} + e_{ij}, \\ \boldsymbol{e}_i & \overset{\text{iid}}{\sim} & ST_{n_i,v}(0, \sigma^2\boldsymbol{I}_{n_i}, \delta\boldsymbol{I}_{n_i}), \\ \beta_{1i} & = & \beta_1 + b_{1i},\ \beta_{2i} = \beta_2 + b_{2i},\ \beta_{3i} = \beta_3 + b_{3i},\ \beta_{4i} = \beta_4 + b_{4i},\ \beta_{5i} = \beta_5 + b_{5i}, \end{array} \tag{12}$$

where $\boldsymbol{\beta} = (\beta_1, \beta_2, \beta_3, \beta_4, \beta_5)^T$ and $\boldsymbol{b}_i = (b_{1i}, b_{2i}, b_{3i}, b_{4i}, b_{5i})^T \overset{\text{iid}}{\sim} N_5(\mathbf{0}, \boldsymbol{\Sigma}_\mathbf{b})$

Model IV: Nonlinear mixed-effects model (two uniexponential):

$$\begin{array}{rcl} y_{ij} & = & \beta_{1i} - [\beta_{2i}\exp(-\beta_{3i}t_{ij}) + \beta_{4i}\exp(-\beta_{5i}t_{ij})]t_{ij} + e_{ij}, \\ \boldsymbol{e}_i & \overset{\text{iid}}{\sim} & ST_{n_i,v}(0, \sigma^2\boldsymbol{I}_{n_i}, \delta\boldsymbol{I}_{n_i}), \\ \beta_{1i} & = & \beta_1 + b_{1i},\ \beta_{2i} = \beta_2 + b_{2i},\ \beta_{3i} = \beta_3 + b_{3i},\ \beta_{4i} = \beta_4 + b_{4i},\ \beta_{5i} = \beta_5 + b_{5i}, \end{array} \tag{13}$$

where $\boldsymbol{\beta} = (\beta_1, \beta_2, \beta_3, \beta_4, \beta_5)^T$ and $\boldsymbol{b}_i = (b_{1i}, b_{2i}, b_{3i}, b_{4i}, b_{5i})^T \overset{\text{iid}}{\sim} N_5(\mathbf{0}, \boldsymbol{\Sigma}_\mathbf{b})$

Model V: Semiparametric mixed-effects model:

$$y_{ij} = \beta_{1i} - v[w(t_{ij}), h_i(t_{ij})]t_{ij} + e_{ij},$$

where $w(t)$ and $h_i(t)$ are unknown nonparametric smooth fixed-effects and random-effects functions, respectively, and $h_i(t)$ are *iid* realizations of a zero-mean stochastic process. Model V is a semiparametric mixed effects model if $w(t)$ and $h_i(t)$ are modeled nonparametrically such as splines or local polynomials. There are several ways to approximate these nonparametric functions. Following the similar approach as Shi et al.(1996), Rice and Wu (2001) and Huang and Dagne (2010), we used natural cubic basis function instead of smoothing splines or kernel methods for two reasons: this method is more straightforward in application and we can select the bases by Akaike information criterion (AIC) or the Bayesian information criterion (BIC) to balance the goodness-of-fit and model complexity. A linear combination of base function can be expressed as:

$$w(t) \approx w_p(t) = \textstyle\sum_{l=0}^{p-1} \mu_l\psi_l(t) = \boldsymbol{\mu}_p\boldsymbol{\Psi}_p(t)^T, \quad h_i(t) \approx h_{iq}(t) = \textstyle\sum_{l=0}^{q-1} \xi_{il}\phi_l(t) = \boldsymbol{\xi}_{iq}\boldsymbol{\Phi}_q(t)^T, \tag{14}$$

where $\boldsymbol{\mu}_p$ and $\boldsymbol{\xi}_{iq}$ ($q \le p$) are the unknown vectors of fixed and random coefficients, respectively. We set $\psi_0 = \phi_0 \equiv 1$ and took the same natural cubic splines in the approximations with $p \le q$, based on the AIC and BIC values, selected as $w(t_{ij}) + h_i(t_{ij}) \approx$

$\mu_0 + \mu_1\psi_1(t_{ij}) + \mu_2\psi_2(t_{ij}) + \xi_{i0}$, where $p = 3$ and $q = 1$. Model V, therefore, can be expressed as,

$$\begin{array}{rcl} y_{ij} & = & \beta_{1i} - [\mu_0\psi_0(t_{ij}) + \mu_1\psi_1(t_{ij}) + \mu_2\psi_2(t_{ij}) + \xi_{i0}]t_{ij} + e_{ij}, \\ \beta_{1i} & = & \beta_1 + b_{1i}, \quad \beta_{2i} = \mu_0 + \xi_{i0}, \quad \beta_3 = \mu_1, \quad \beta_4 = \mu_2, \end{array} \tag{15}$$

where $\boldsymbol{\beta} = (\beta_1,\ \mu_0,\ \mu_1,\ \mu_2)^T$ and $\boldsymbol{b}_i = (b_{1i}, \xi_{i0})^T \overset{\text{iid}}{\sim} N_2(\mathbf{0}, \mathbf{\Sigma_b})$.

In each of the five models above, besides the ST distribution assumption, the model random error can also be assumed to follow other more specific distributions as normal, SN and Student-*t*. We used several criteria to check the model fit by applying the models on the data mentioned above.

We first used deviance information criterion (DIC) (Spiegelhalter, 2002) to compare models. Same as AIC and BIC, the smaller the value of DIC, the better of the model fit. DIC is not intended for identification of the 'correct' model, but rather merely as a way to compare a collection of alternative formulations.

Because model comparisons are critical for our study, besides DIC, we also compared the values of expected predictive deviance (EPD) and residual sum of squares (RSS) that are obtained from each model. EPD is formulated by $EPD = E\{\mathbf{\Sigma}_{i,j}(y_{rep,ij} - y_{obs,ij})^2\}$, 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 $\boldsymbol{\theta}$ (Gelman et al., 2003). RSS is given by $\mathbf{\Sigma}_{i,j}(y_{obs,ij} - y_{fitted,ij})^2$. The smaller the value of DIC, EPD and RSS, the better fit of the model to the data. Besides these statistical criteria, two types of plots, Quantile-Quantile (Q-Q) plot and observed values vs. fitted values, were also reported to give a visualized goodness-of-fit in the model comparisons.

In the Bayesian inferential approach, we also need to specify values of the hyper-parameters at the population level. Weakly informative prior distributions are taken for all the parameters: (i) for each component of fixed-effects vector of $\boldsymbol{\beta}$, the prior was assumed to follow independent normal distribution as $N(0, 100)$; (ii) for the scale parameter σ^2, we assumed a limiting non-informative inverse gamma prior distribution as $IG(0.01, 0.01)$, therefore, the mean is 1 and variance is 100; (iii) the prior for the variance-covariance matrix for the random-effect $\mathbf{\Sigma}_b$ was taken to be inverse Wishart distribution as $IW(\mathbf{\Omega}, v)$, the degree of freedom, $v = 5$, and $\mathbf{\Omega}$ is diagonal matrix with diagonal elements being 0.01; (iv) for the skewness parameter δ, we chose normal distribution; (v) the degree of freedom ν followed truncated exponential distribution with $\nu_0 = 0.5$

The MCMC sampler was implemented using WinBUGS package (Lunn et al., 2000). The posterior means and quantiles were drawn after the collecting the final MCMC samples. We used one long chain. Convergence, which refers the algorithm has reached its equilibrium target distribution, was closely watched by using the standard tools within WinBUGS such as trace plots, the MC error and depicting the evolution of the ergodic means of a quantity over the number of iterations. After an initial 100,000 burn-in iterations, every 50*th* MCMC sample was retained from the next 200,000. Thus, we obtained 4,000 samples of targeted posterior distribution of the unknown parameters for statistical inference.

We carry out model comparisons in two steps as follows. Step 1: in Section 4.2, we determine when the model errors are assumed to have an ST distribution, among the five models presented in Section 4.1, which one has the best fit. Step 2: in Section 4.3, because normal, SN and Student-*t* distributions are all a special case of an ST distribution, focusing on the best model selected from Step 1, we compare the results based on random errors with the normal, SN, Student-*t* and ST distributions. The model comparisons are carried on A5055 data and confirmed by A398 data. Section 4.4 presents the results based on the best model selected.

4.2 Comparison of Five Models under an ST Distribution

We should not directly compare the parameter values obtained from the models that have different components. However, because the models were applied to the same HIV viral load data, we could use DIC, EPD and RSS to find out which model had the best fit. Table 1 indicates that Model IV has the best fit. Among all of the data sets: ACTG5055, the two randomly selected samples from ACTG398 and ACTG398 that includes every subject, Model IV constantly has the lowest DIC, EPD and RSS. For example, in ACTG5055, the DIC value for Model IV is 21.7, while the DIC values are 1192.0, 401.1, 669.2, and 1015.7 in Models I, II, III and V, respectively.

Table 1: DIC, EPD and RSS among the five models, random error is assumed to follow an ST distribution.

Data set		Model I	Model II	Model III	Model IV	Model V
ACTG5055:	DIC	1192.0	401.1	669.2	**21.7**	1015.7
All subjects	EPD	0.33	0.12	0.15	**0.05**	0.20
(n=44)	RSS	59.2	20.7	27.1	**8.4**	35.3
ACTG398:	DIC	1239.4	1122.6	1162.9	**576.9**	1200.5
Sample 1	EPD	3.17	3.48	3.63	**1.04**	3.04
(n=44)	RSS	396	436	454	**134**	381
ACTG398:	DIC	2264.9	1947.5	2855.3	**1757.7**	2204.5
Sample 2	EPD	0.44	0.51	3.45	**0.33**	0.39
(n=100)	RSS	130.6	158.3	1083	**99.39**	116.7
ACTG398:	DIC	12900.6	10891.2	10763.4	**8819.3**	14217
All subjects	EPD	1.49	1.42	1.31	**0.93**	1.84
(n=481)	RSS	2665	2185	2059	**1470**	2703

4.3 Comparison of the Best Model IV with Four Distributions for Random Error

For Model IV, we further investigated how different distributions of random error would affect the model fit and the DIC values are shown in Table 2 below. The model with ST (ACTG5055, two samples of ACTG398) has the lowest DIC.

Table 2: DIC values based on For Model IV with different distribution assumptions.

Distribution	ACTG5055 (n=44)	ACTG398 (n=44)	ACTG398 (n=100)
Normal	1133.3	961.4	2158.5
SN	222.8	727.3	1937.2
Student-*t*	1004.9	949.8	2144.3
ST	**21.7**	**576.9**	**1757.7**

We applied Model IV on ACTG5055 to further compare the estimation results obtained from different distributions. The posterior mean (PM), the corresponding standard deviation (SD) and 95% credible interval (CI) for fixed-effects parameters are presented in Table 3. We found that: (i) with the exception β_5 based on the normal distribution, all of the other estimates were significant because the 95% CIs don't include zero; (ii) for the variance σ^2, the estimated value based on the SN (0.05) and ST (0.01) models were much smaller than the model with normal (1.15) or Student-*t* assumption (0.38); (iii) among all of the parameters estimated, the SD obtained from ST were the smallest; (iv) the estimates were similar between normal and Student-*t* model, but they were substantially different to those obtained from SN or ST model. For example, β_2 based on normal and Student-*t* models was 34.67 and 38.10, respectively, while it was 24.53 and 27.89 in SN and ST models, respectively;

(v) the skewness parameter δ was significantly positive in SN and ST models, confirming the positive skewness of the viral load in natural $\log$ scale as shown in Figure 1(a); (vi) compared to the model with normal or Student-*t* random error, the models with an SN or ST fit the data better. For example, in ACTG5055, for DIC values, 1133.3 (normal) vs. 222.8 (SN), 1004.9 (Student-*t*) vs. 21.7 (ST), it indicates that consideration of a departure from normality will improve the model fitting.

Table 3: A summary of the estimated posterior mean (PM) of interested population (fixed-effects), precision and skewness parameters, the corresponding standard deviation (SD) and lower limit (L_{CI}) and upper limit (U_{CI}) of 95 % equal-trial credible interval (CI) from Model IV with normal, SN, Student-*t* and ST distributions of random error (based on ACTG5055 data).

Model IV		β_1	β_2	β_3	β_4	β_5	σ^2	δ	DIC	EPD	RSS
Normal	PM	8.30	34.67	5.95	6.77	0.38	1.15	–	1133.3	2.29	1133.3
	L_{CI}	7.81	26.86	4.01	2.65	-0.40	0.96	–			
	U_{CI}	8.77	43.04	8.54	12.92	1.01	1.37	–			
	SD	0.26	4.15	1.26	2.52	0.35	0.11	–			
SN	PM	6.69	24.53	6.38	13.49	1.55	0.05	2.22	222.8	0.10	17.5
	L_{CI}	6.03	12.66	4.10	7.10	0.89	0.01	1.97			
	U_{CI}	7.40	35.03	11.22	20.19	2.15	0.16	2.51			
	SD	0.35	5.69	1.72	3.85	0.37	0.04	0.14			
Student-*t*	PM	8.33	38.10	9.41	11.97	0.94	0.38	–	1004.9	2.19	420.8
	L_{CI}	7.84	30.31	7.34	8.99	0.66	0.28	–			
	U_{CI}	8.83	46.61	11.64	15.59	1.24	0.50	–			
	SD	0.25	4.11	1.13	1.66	0.15	0.06	–			
ST	PM	7.18	27.89	7.19	13.69	1.73	0.01	1.17	**21.7**	**0.05**	**8.4**
	L_{CI}	6.66	21.53	6.34	11.72	1.55	0.00	0.99			
	U_{CI}	7.67	33.93	8.63	16.22	1.98	0.04	1.34			
	SD	0.25	3.26	0.66	1.28	0.13	0.01	0.08			

Several diagnostic plots for goodness-of-fit are also applied. Firstly, we select three representative subjects corresponding to the three typical patterns trajectories displayed in Figure 1(a). These three trajectory patterns represent (i) rapid decline followed by slow decrease, (ii) rapid decline followed by rebound and then decrease, and (iii) decline followed by rebound. The individual estimates of viral load trajectories are shown in Figure 3. The following findings are observed: (i) the estimated individual trajectories obtained from SN and ST fit the originally observed data much closer than those got from the model where the random error is assumed to be normal or Student-*t*; (ii) the average SD obtained from ST is the smallest, which is 0.15, while the mean of SD for the individual estimation got from SN, N and Student-*t* is 0.22, 0.52 and 0.46, respectively. Note that the lack of smoothness in SN and ST model estimates of individual trajectories is understandable since a random component $\boldsymbol{w}_i$ was incorporated in the expected function (see equation (7) for details) according to the stochastic representation feature of the SN and ST distributions for "chasing the data" to this extent.

We also applied two diagnostic plots: plot of the observed values versus the fitted values and Q-Q plot (not shown here). The findings agree with that from DIC criterion: the models with SN and ST distributions provided better fit to the observed data than the ones with normal or Student-*t* distribution. Based on the results from DIC, EPD, RSS and the diagnostic plots, we conclude that Model IV with an ST distribution fits the data better than the other combinations among the models with a different time-varying decay rate function and/or distribution assumption of random errors.

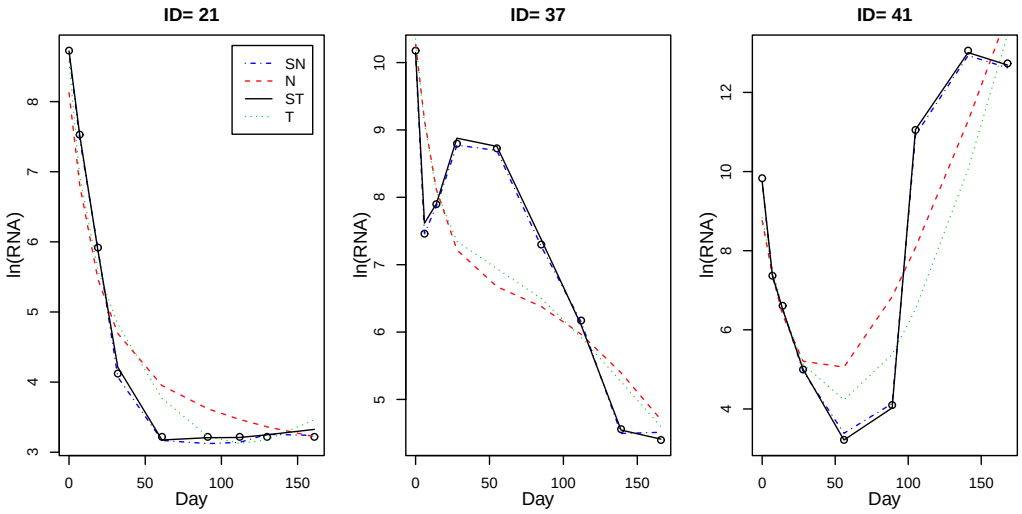

Figure 3: Individual estimates of viral load trajectories for three representative patients based on Model IV with normal, Student-*t*, SN and ST distributions. The observed values are indicated by circle.

4.4 Results Based on Model IV with an ST Distribution

Based on Model IV with an ST distribution, the estimated population decay rate function based on ACTG5055 data is

$$\hat{\lambda}(t) = 27.89\exp(-7.9t) + 13.69\exp(-1.73t).$$

Because the estimated $\hat{\lambda}(t)$ is always positive, the population viral load would always decrease in this specific HIV/AIDS data set.

The individual time-varying decay rate function is given by,

$$\hat{\lambda}(t_{ij}) = \hat{\beta}_{i2}\exp(-\hat{\beta}_{i3}t_{ij}) + \hat{\beta}_{i4}\exp(-\hat{\beta}_{i5}t_{ij})$$

where the individual estimated decay rate $\hat{\lambda}(t_{ij})$ is considered to be dependent on both subjects and time points. We found that the individual decay rate at initial treatment, $\hat{\lambda}(t_{i1})$, where $t_{i1} = 0$, was positively correlated with baseline viral load (Spearman correlation coefficient r = 0.769, $p < 0.0001$) and negatively associated with baseline CD4 cell count (r = -0.447, $p = 0.0025$). Overall, the individual decay rate, $\hat{\lambda}(t_{ij})$, was positively associated with viral load ($p < 0.0001$) and negatively associated with CD4 cell count ($p < 0.0001$).

Because 30 $\sim$ 60% (Havlir et al., 2000) of patients eventually will have viral rebound, it is important to have a model that can reasonably predict this type of treatment failure in the long term. Following Wu et al.(2008), we defined rebound as, comparing with the HIV-1 viral load (natural log transformed) from the previous measurement, if there was $\geq$ 1.15 increase at one time point or $\geq$ 0.46 increase at two or more consecutive time points. In ACTG5055, there were 11 (26.2%) subjects had rebound. There was no significant difference in the baseline viral load (natural $\log$ (RNA)) between the rebound and no rebound group (median was 9.18 and 8.78/mL, respectively, p = 0.8610), while the median of baseline CD4 cell count intended to be higher in the no rebound group than in the rebound group (285 vs. 253/mL, respectively, p = 0.1169).

The trend of the changes in decay rates during the treatment was different between the rebound and no rebound group (Figure 4). For example, every individual decay rate was positive in the no rebound group, while some individual decay rates in the rebound group

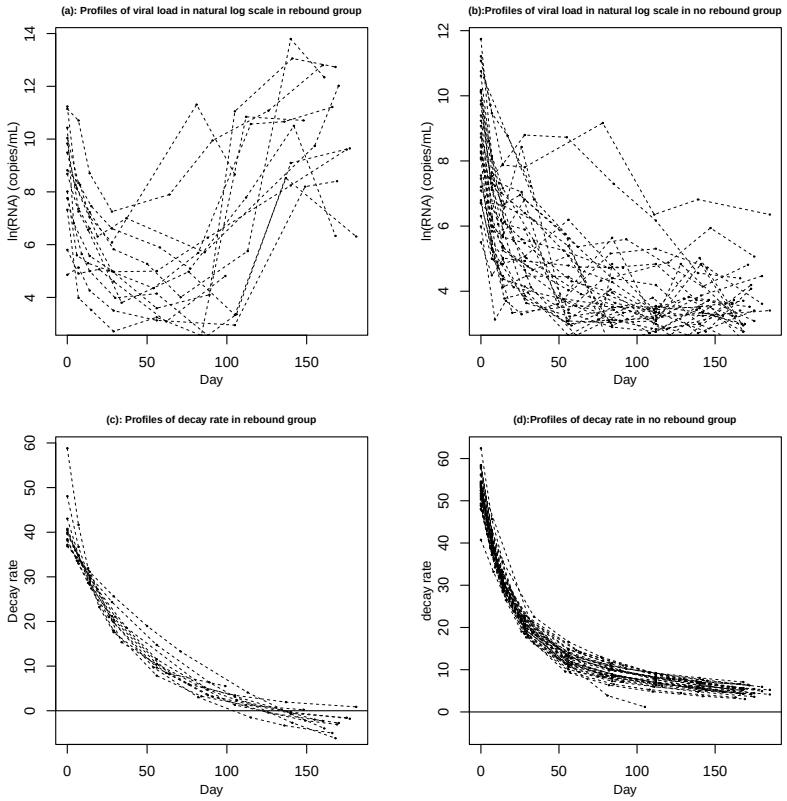

Figure 4: Profiles of viral load in natural $\log$ scale and decay rate in rebound and no rebound group.

became negative, especially after the 3*rd* month of the treatment, which corresponded to the viral load rebound.

Based on the results of Model IV under the ST distribution in ACTG5055 data, we also found that: (i) overall, the average value of individual decay rates, $\hat{\lambda}(t_{ij})$, was bigger in the no rebound group (14.97) than in the rebound group (12.93); (ii) the initial individual decay rates, $\hat{\lambda}(t_{i1})$, were significantly bigger in the no rebound group than in the rebound group (mean is 53.16 and 40.95, respectively); (iii) $\hat{\lambda}(t_{i1})$ was significantly associated with the rebound status in the long term (OR = 0.703, 95% CI is 0.580 – 0.853, p = 0.0003) and this association was still significant even after controlling the baseline viral load and CD4 cell count (OR = 0.717, 95% CI is 0.588 – 0.875, p = 0.0010). (iv) the average individual decay rate at the last visit ($\hat{\lambda}(t_{in_i})$) among the no rebound subjects was 4.67, while it was -2.28 in the rebound group which indicates the viral load actually increasing instead of decreasing in this group. Among these findings, the most interesting is the associated relationship between initial individual decay rate ($\hat{\lambda}(t_{i1})$) and rebound status: the results indicated that the odds of rebound decreased about 30% with each 1 unit increased in the decay rate. This may be helpful for physicians to predict the long-term results based on the information at early stage of the disease.

5. Concluding Discussion

With an ST distribution assumption for model random errors to account skewness observed in viral load responses, we compared five commonly used mixed-effects models in HIV dynamics via the Bayesian approach. We also investigated the impact of the four distributions in the skew-elliptical family on the model fitting. The results indicate that with

the ST distribution, there is potential gain of efficiency and accuracy in estimating certain parameters when the normality assumption does not apply to the data. The skew-elliptical modeling via the Bayesian approach proposed in this study can be easily carried out via the WinBUGS package. Because the proposed model is quite general in theory and accessible to the existing software, it will allow statisticians to apply this method in other fields.

After finding the best fitting model, we estimated the relationship between the individual viral decay rate and some clinical important variables. The initial individual decay rate was positively correlated with the baseline viral load and negatively associated with baseline CD4 cell count. We also found that, overall, the average individual decay rate was lower in the rebound group than in the no rebound group. A more interesting finding is the significant association between the initial individual decay rate and the rebound status in the long-term, even after controlling for the baseline viral load and CD4 cell count. This finding is clinically important because it may enable physicians to predict the long-term outcome based on the estimated decay rate at an early stage.

Using the model with a time-varying decay rate function has some advantages over the biphasical models. (i) In the biphasical models, the association between the first decay rate and baseline viral load could be positive (Notermans et al., 1998; Wu et al., 2004) or negative (Wu et al., 1999); no significant association was found between the rebound and the first decay rate either (Wu et al., 2008); (ii) although the second decay rate in the biphasical models is supposed to be associated with long-term treatment status such as rebound (Ding and Wu, 1999; Wu et al.,2005), no significant association was found between the second decay rate and the low-level viral replication in long term (Sedaghat et al., 2008; Wu et al., 2003).

This paper has some limitations. Usually, covariates are included in the mixed-effects model to control within- and between-subject variation, and CD4 cell count is a commonly used covariate in HIV dynamic models. However, in order to use the original proposed models in the comparisons, we did not include covariates such as CD4 cell count or demographic information. For the viral load, the values below the detectable level are usually considered as inaccurate. Instead of treating these values as censored, we computed them by half the value of the detectable level. Furthermore, the issue of missing values is not considered in this study either.

In conclusion, the skewness parameter in the model with SN or ST distribution assumption is significantly positive, which confirms the positive skewness observed in the viral load data even after natural log transformed. The model fit is the best in the model with skewed (SN or ST) distribution. Because estimated parameters can be considerably different between the models with skewed distribution and normal or Student-*t* distribution, it is important to account for skewness in the model when data exhibits noticeable skewness. Different models may yield different conclusions about the relationship between the individual decay rate with viral load, CD4 cell count and the rebound status in HIV dynamics, therefore, it is also critical to choose a reasonable model that can balance between complexity and utility.

ACKNOWLEDGEMENTS

This research was partially supported by USF Proposal Enhancement Grant to Y. Huang.

References

Acosta, E.P., Wu, H., Walawander, A., Eron, J., Pettinelli, C., Yu, S.,..., Gerber, J.G. (2004). Comparison of two indinavir/ritonavir regimens in treatment-experienced HIV-infected individuals.*Journal of Acquired Immune Deficiency Syndromes*, 37, 1358–1366.

Beal, S.L., Sheiner, L.B. (1982). Estimating population kinetics. *Critical Review in Biomedical*

Engineering, 8, 192–222.

Commenges, D., Jolly, D., Drylewicz, J., Putter, H., Thiebaut, R. (2011). Inference in HIV dynamics models via hierarchical likelihood. *Biometrics* 55(1):446–456.

Dagne, G., Huang, Y. (2012). Mixed-effects Tobit joint models for longitudinal data with skewness, detection limits and measurement errors. *Journal of Probability and Statistics*, Volume 2012 (2012), Article ID 614102.

Davidian, M., Giltinan, D.M. (1995). *Nonlinear Models for Repeated Measurement Data*. London-Chapman & Hall.

Ding, A.A., Wu, H. (1999). Relationships between antiviral treatment effects and biphasic viral decay rates in modeling HIV dynamics. *Mathematical Biosciences*, 160, 63–82.

Gelman, A., Carline, J.B., Stern, H.S., Rubin, D.B. (2003). *Bayesian data analysis*. Chapman & Hall/CRC. London.

Gardner R.C. (2001). *Pyschological statistics using SPSS for Windows.* McGraw-Hill.

Grossman, Z., Polis, M., Feinberg, M., Grossman, Z., Levi, I., Jankelevich, S., Yarchoan, R., Boon, J., de Wolf, F., Lange, J.M.A., Goudsmit, J., Dimitrov, D.S., Paul, W.E. (1999). Ongoing HIV dissemination during HAART. *Nature Medicine,* 5(10), 1099–1104.

Havlir, D.V., Hellmann, N.S., Petropoulos, C.J., Whitcomb, J.M., Collier, A.C., Hirsch, M.S.,..., Richman, D.D. (2000). Drug susceptibility in HIV infection after viral rebound in patients receiving Indinavir-containing regimens. *Journal of the American Medical Association,* 283, 229–234.

Helsel, D.R. (1990). Less than obvious-statistical treatment of data below the detection limit. *Environmental Science and Technology*, 24(12), 1766–1774.

Ho, D.D., Neumann, A,U., Perelson, A.S., Chen, W., Leonard, J.M., Markowitz, M. (1995). Rapid turnover of plasma virions and CD4 lymphocytes in HIV-1 infection. *Nature,* 373, 123–126.

Huang, Y., Liu, D., Wu, H. (2006). Hierarchical Bayesian Methods for Estimation of Parameters in a Longitudinal HIV Dynamic System. *Biometrics,* 62(2), 413–423.

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

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

Kuhn, E., Lavielle, M. (2005). Maximum likelihood estimation n nonlinear mixed effects models. *Computational Statistics and Data Analysis*, 49, 1020–1038.

Lavielle, M., Samson, A., Karine-Fermin, A., Mentre E. (2011). Maximum likelihood estimation of long-term HIV dynamic models and antiviral response. *Biometrics,* 67(1), 250–259.

Lindstrom, M., Bates, D. (1990). Nonlinear mixed effects models for repeated measures data. *Biometrics,* 46, 673–687.

Liu, W., Wu, L. (2007). Simultaneous inference for semiparametric nonlinear mixed-effects models with covariate measurement errors and missing responses. *Biometric,* 63, 342–350.

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

Maldarelli, F., Palmer, S., King, M.S., Wiegand, A., Polis, M.A., Mican, J., Kovacs, J.A., ..., Mellors, J.M. (2007). ART suppresses plasma HIV-1 RNA to a stable set point predicted by pretherapy viremia. *PloS Clinical Trials,* 3(4), e46–e87.

Notermans, D.W., Goudsmit, J., Danner, S.A., De-Wolf, F., Perelson, A.S., Mittler, J. (1998). Rate of HIV-1 decline following antiretroviral therapy is related to viral load at baseline and drug regimen. *AIDS,* 12, 1483–1490.

Perelson, A.S., Essunger, P., Cao, Y., Vesane, M., Hurley, A., Saksela, K., Markowltz, M., Ho, D.D. (1997). Decay characteristics of HIV-1-infected compartments during combination therapy. *Nature,* 387, 188–191.

Pfister, M., Labbe, L., Hammer, S.M., Mellors, H., Bennett, K.K., Rosenkranz, S., Sheiner, L.B. and the AIDS clinical trial group protocol 398 investigators. Population pharmacokinetics and pharmacodynamics of Efavirenz, Nelfinavir and Indinavir: adult AIDS clinical trial group study 398. *Antimicrobial Agents and Chemotherapy,* 47(1), 130–137.

Putter, H., Heisterkamp, S.H., Lange, J.M.A., Wolf, F. (2002) A Bayesian approach to parameter estimation in HIV dynamic models. *Statistics in Medicine*, 21, 2199–2214.

Rice, J.A., and Wu, C.O. (2001). Nonparametric mixed-effects models for unequally sampled noisy curves. *Biometrics*, 57, 253–259.

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

Shi, M., Weiss, R.E., Taylor, JMG. (1996). An analysis of pediatric CD4+ counts for acquired immune deficiency syndrome using flexible random curves. *Applied Statistics*, 45, 151163.

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

Verbeke, G. Lesaffre, E. (1996). A linear mixed-effects model with heterogeneity in random-effects population. *Journal of the American Statistical Association*, 91, 217–221.

Wei, X., Ghosh, S.K., Taylor, M.E., Johnson, V.A., Emini, E.A., Deutsch, P.D., et al. (1995). Viral dynamics in human immunodeficiency virus type 1 infection. *Nature*, 373(12), 117–122.

Wu, H., Kuritzkes, D.R., McClernon, D.R., Kessler, H., Connick, E., Landay, A., , Lederman, M, M. (1999). Characterization of viral dynamics in human immunodeficiency virus type 1-patients treated with combination antiretroviral therapy: relationship to host factors, cellular restoration, and virologic end points. *The Journal of Infectious Diseases*, 179, 799-807.

Wu, H., Zhang, J-T. (2002). The study of long-term HIV dynamics using semi-parametric non-linear mixed-effects models. *Statistics in Medicine*, 21, 3655–3675.

Wu, H., Zhao, C., Liang, H. (2004). Comparison of linear, nonlinear and semiparametric mixed-effects models for estimating HIV dynamics parameters. *Biometrical Journal*, 6, 233–245.

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

Wu, H., Huang, Y,, Acosta, E.P., Rosenkranzm S.L., Kuritzkesm D.R., Eronm J.J., Perelson, A.S., Gerber, J.G. (2005). Modeling long-term HIV dynamics and antiretroviral response. *Journal of AIDS*. 39: 272–283.

Wu, L., Hu, X.J., Wu, H. (2008). Joint inference for nonlinear mixed-effects models and time to event at the presence of missing data. *Biostatistics*, 9(2), 308–320.

Zhang, J., WU, H. (2011). Modeling HIV dynamics using unified mixed-effects models. *American Journal of Mathematical and Management Sciences*, 30(2), 83–111.