REVSTAT – Statistical Journal
Volume 13, Number 2, June 2015, 169–191

# A BAYESIAN APPROACH FOR JOINT MODELING OF SKEW-NORMAL LONGITUDINAL MEASUREMENTS AND TIME TO EVENT DATA

Authors: Taban Baghfalaki
– Department of Statistics, Faculty of Mathematical Sciences, Tardit Modares University, Tehran, Iran
t.baghfalaki@modares.ac.ir

Mojtaba Ganjali
– Department of Statistics, Faculty of Mathematical Sciences, Shahid Beheshti University, Tehran, Iran
m-ganjali@sbu.ac.ir

Received: April 2013 Revised: December 2013 Accepted: February 2014

Abstract:

- Joint modeling of longitudinal measurements and survival time has an important role in analyzing medical data sets. For example, in HIV data sets, a biological marker such as CD4 count measurements is considered as a predictor of survival. Usually, longitudinal responses of these studies are severely skew. An ordinary method for reducing the skewness is the use of square root or logarithm transformations of responses. In most of the HIV data sets, because of high rate of missingness, skewness is remained even after using the transformations. Therefore, a general form of distributions for considering skewness in the model should be used. In this paper, we have used multivariate skew-normal distribution to allow a flexible model for considering non-symmetrically of the responses. We have used a skew-normal mixed effect model for longitudinal measurements and a Cox proportional hazard model for time to event variable. These two models share some random effects. A Bayesian approach using Markov chain Monte Carlo is adopted for parameter estimation. Some simulation studies are performed to investigate the performance of the proposed method. Also, the method is illustrated using a real HIV data set. In these data, longitudinal outcomes are skew and death is considered as the event of interest. Different model structures are developed for analyzing this data set, where model selection is performed using some Bayesian criteria.

Key-Words:

- *Bayesian approach; Cox proportional model; joint modeling; longitudinal data; skew-normal distribution; time to event data.*

AMS Subject Classification:

- 62H99, 62J99.

## 1. INTRODUCTION

In most of the HIV and cancer studies a longitudinal biological marker such as CD4 count or immune response can be an important predictor of survival. In these studies a time to an event may also be a variable of interest. Patients are monitored longitudinally and some longitudinal measurements are gathered until the interest event occur. Often the longitudinal outcomes and time to event of interest are analyzed jointly using joint modeling of longitudinal and time to event data.

Joint modeling of longitudinal measurements and time to event data has been studied by DeGruttola and Tu (1994), Tsiatis *et al.* (1995) and Wulfsohn and Tsiatis (1997). Henderson *et al.* (2000) and Hashemi *et al.* (2003) discussed joint modeling of longitudinal measurements and event time data using latent class of Gussian process, Tsiatis and Davidian (2001), Yu *et al.* (2004) and Sousa (2011) provide reviews of this joint modeling. Tseng *et al.* (2005) used accelerated failure time model for joint modeling of longitudinal and survival data and applied Monte Carlo EM approach to estimate unknown parameters. Also, Diggle *et al.* (2008) discussed different approaches to estimate unknown parameters of joint modeling of longitudinal measurements and event time data and then applied a fully parametric approach to modeling Schizophrenic patients data set. Joint Modeling of longitudinal measurements and time to event data at the presence of informative dropout in a HIV study was discussed by Wu *et al.* (2008). They considered an additional missingness mechanism for missing values. Rizopoulos (2010) presented the R package "*JM*" that can be used to fit the joint modeling of longitudinal measurements and survival data. Also, Guo and Carlin (2004) discussed the implementation of the joint models in SAS and WinBUGS under normal distributional assumption.

In the above mentioned references, usually a mixed effect model with normality or other symmetric distributional assumption is used for longitudinal part of the model. However, in the most of such studies longitudinal measurements are severely skew or have some outliers. For the latter problem, some models, which are robust in the existence of outliers, have been considered by Lachos *et al.* (2010) and Bandyopadhyay *et al.* (2010). They have used skew-normal/independent distribution assumptions in linear mixed effect models. However, the problem of skewness, to our best of knowledge has not yet been considered in joint modeling of longitudinal and time to event data. The idea is that a parametric skew distribution may be considered for this regard.

Skew-normal (SN) family is an important class of non-symmetric distribution for analyzing abnormal data set. The distribution includes normal one as a special case. The first version of the distribution which is in the univariate form is introduced by Azzalini (1985). More discussion about univariate skew-normal

distribution can be found in Azzalini (1986) and Henez (1986). Generalizations to the multivariate case are given in Azzalini and Dalla-Valle (1996), Azzalini and Capitanio (1999), Branco and Dey (2001) and Sahu *et al.* (2003). For examples, some applications of skew-normal in regression model can be found in Lachos *et al.* (2007), Cancho *et al.* (2010) and Arellano-Valle *et al.* (2005b). Multivariate skew-normal mixed effect model have discussed by Arellano-Valle *et al.* (2005a) and Lin and Lee (2008), also, discussion about multivariate skew-normal with incomplete data can be found in Lin *et al.* (2009) and Baghfalaki and Ganjali (2011, 2012). Recently, there are some applications of skew-normal distribution for analysing HIV data. For examples: Ghosh and Branco (2007) develop a Bayesian approach to bivariate random effect model with application to HIV studies. Huang and Dagne (2011) used skew-normal distribution in a Bayesian approach to joint modeling mixed effect and measurement error for a HIV study. Huang and Dagne (2010) developed a Bayesian nonlinear mixed effect model with skew-normal random effect and within subject errors for providing a better fit to HIV data set, Huang *et al.* (2011a) suggested linear, nonlinear and semiparametric mixed-effect model using skew-normal distribution with measurement error in covariates for analyzing an AIDS data set and Lachos *et al.* (2011) developed a Bayesian framework for analyzing censored data using linear or non-linear model under skew-normal/independent distributional assumption with application in HIV studies. Huang *et al.* (2011b) used skew-normal distribution for joint modeling of CD4 process and time to increase CD4/CD8 ratio. They used a mixed effect model for analyzing the longitudinal measurements alongside a log-normal model for analyzing event time data.

In this paper, we have discussed Bayesian joint modeling of longitudinal and survival data when skewness exists in the data set. We have used multivariate skew-normal distribution introduced by Sahu *et al.* (2003) for considering skewness of the data. Implementation of the Bayesian approach using this form is easier than other forms of skew-normal distributions. A non-ignorable missingness mechanism is considered for missingness. Also, a skew-normal mixed effect model and a Cox proportional hazard model (as a semiparametric model) with step baseline hazard in a frailty model structure are considered for the joint modeling. We have performed some simulation studies to investigate the performance of the proposed method with different sample sizes and different rates of drop out. We have used the proposed method for analyzing a HIV study, where CD4 count measurements are longitudinal measurements and time to death is considered as the interest event. The aim of the study was to compare the efficacy and safety of two alternative antiretroviral drugs, namely didanosine (ddI) and zalcitabine (ddC). In this data set CD4 count measurement is a skew variable which is gathered along side with the time to the event of interest. We have used the proposed model and pure normal model for analyzing the data set. The results of using these distributional assumptions are compared using some criteria; also some influential observations are detected using Kullback–Leibler divergence.

The rest of this paper is organized as follows. In Section 2, we introduce multivariate skew-normal distribution which we will use in this paper. Section 3 includes the model and notations of the paper. In that Section, the model for longitudinal and survival part is described separately. In Section 4 Bayesian approach of joint modeling of longitudinal measurements and event time data using multivariate skew-normal distribution is discussed. Section 5 includes some simulation studies for investigating the proposed model. In Section 6, we apply the proposed approach to the HIV data set and finally concluding remarks are given in Section 7.

## 2. MULTIVARIATE SKEW-NORMAL DISTRIBUTION

Multivariate skew-normal distributions have different forms, some of these distributions have been introduced by: Azzalini and Dalla-Valle (1996), Azzalini and Capitanio (1999), Arellano-Valle and Genton (2005) and Arellano-Valle *et al.* (2005b). One of the commonly used multivariate skew-normal distributions, in Bayesian context, is introduced by Sahu *et al.* (2003). In this section, we review this form.

Let $\phi_k(\boldsymbol{y}|\boldsymbol{\mu},\boldsymbol{\Sigma})$ and $\Phi_k(\boldsymbol{y}|\boldsymbol{\mu},\boldsymbol{\Sigma})$ be the probability density function and cumulative density function of the $N_k(\boldsymbol{\mu},\boldsymbol{\Sigma})$ evaluated at $\boldsymbol{y}$, respectively. A $k$-dimensional random vector $\boldsymbol{Y}$ follows a $k$-variate skew-normal distribution with location vector $\boldsymbol{\mu} \in R^k$, $k\times k$ positive definite scale matrix $\boldsymbol{\Sigma}$ and $k\times k$ skewness matrix $\boldsymbol{\Delta} = \text{diag}(\delta_1, ..., \delta_k)$, where $\text{diag}(a_1, ..., a_k)$ denotes a diagonal matrix with elements $a_1, ..., a_k$, if its density function is given by:

$$
\begin{aligned}
f\big(\boldsymbol{y}|\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{\Delta}\big) \;=\; & 2^k\,\phi_k\big(\boldsymbol{y}|\boldsymbol{\mu},\boldsymbol{\Sigma}+\boldsymbol{\Delta}\boldsymbol{\Delta}'\big) \\
& \times\; \Phi_k\Big(\boldsymbol{\Delta}'\big(\boldsymbol{\Sigma}+\boldsymbol{\Delta}\boldsymbol{\Delta}'\big)^{-1}(\boldsymbol{y}-\boldsymbol{\mu})\,\big|\,0\,,\;\big(\boldsymbol{I}_k+\boldsymbol{\Delta}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\Delta}\big)^{-1}\Big)\,.
\end{aligned}
\tag{2.1}
$$

We denote this by $\boldsymbol{Y} \sim SN_k(\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{\Delta})$. The mean and covariance matrix of $\boldsymbol{Y}$ are given, respectively, by:

$$
E[\boldsymbol{Y}] = \boldsymbol{\mu} + \sqrt{\frac{2}{\pi}}\,\boldsymbol{\delta} \quad \text{and} \quad \text{cov}(\boldsymbol{Y}) = \boldsymbol{\Sigma} + \left(1-\frac{2}{\pi}\right)\boldsymbol{\Delta}^2\,,
$$

where $\boldsymbol{\delta} = (\delta_1, ..., \delta_k)'$. The use of the following proposition which is called stochastic representation, makes it possible to generate a sample from skew-normal distribution using available software.

**Proposition 2.1.** *Let* $\boldsymbol{Y} \sim SN_k(\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{\Delta})$, *then* $\boldsymbol{Y} \overset{\text{d}}{=} \boldsymbol{\Delta}|\boldsymbol{X}_0| + \boldsymbol{X}_1$, $\boldsymbol{X}_0 \sim N_k(0,\boldsymbol{I}_k)$, $\boldsymbol{X}_1 \sim N_k(\boldsymbol{\mu},\boldsymbol{\Sigma})$ *and* $\boldsymbol{X}_0$ *and* $\boldsymbol{X}_1$ *are independent. The notation* $\overset{\text{d}}{=}$ *means "distributed as".* $|\boldsymbol{X}_0|$ *is the vector of the absolute values of each component of* $\boldsymbol{X}_0$.

For proof of this proposition see Sahu *et al.* (2003) and Arellano-Valle *et al.* (2007).

This proposition is used for obtaining a hierarchical set-up, that is $\boldsymbol{Y}|\boldsymbol{U} = \boldsymbol{u} \sim N_k(\boldsymbol{\mu} + \boldsymbol{\Delta u}, \boldsymbol{\Sigma})$ and $\boldsymbol{U} \sim N_k(\boldsymbol{0}, \boldsymbol{I}_k)$. This hierarchical set-up has been used broadly in Bayesian context of skew-normal model.

## 3. NOTATION AND SEPARATE MODELS

In this section the notations and models for each part of JM are described. The next section will discuss joint modeling (JM) of longitudinal responses and time to event data with skew-normal distribution assumption for responses.

For longitudinal model, let $y_i(s)$ denote the value of longitudinal outcome at time point $s$ for the $i^{th}$ individual where the observed times are $s_{ij}$, $i = 1, 2, ..., n$, $j = 1, 2, ..., n_i$. In this case we shall write $y_i(s_{ij}) = y_{ij}$. We consider the following linear mixed effect model for describing longitudinal outcome:

$$y_{ij} = \boldsymbol{x}'_{1i}(s_{ij})\boldsymbol{\beta}_1 + \boldsymbol{z}'_{1i}(s_{ij})\boldsymbol{b}_{1i} + \boldsymbol{\varepsilon}_{ij} \ , \qquad i = 1, 2, ..., n \ , \quad j = 1, 2, ..., n_i \ ,$$

where components of $\boldsymbol{\varepsilon}_i = (\varepsilon_{i1}, ..., \varepsilon_{in_i})'$ are measurement errors, $\boldsymbol{\beta}_1 = (\beta_{11}, ..., \beta_{1p_1})'$ is a $p_1$-dimensional vector of longitudinal fixed-effect parameters. $\boldsymbol{b}_{1i} = (b_{1i1}, ..., b_{1iq_1})'$ is a $q_1$-dimensional vector of random effects and is independent of $\boldsymbol{\varepsilon}_i$. $\boldsymbol{x}_1$ and $\boldsymbol{z}_1$ are $p_1$-dimensional and $q_1$-dimensional vectors of explanatory variables, respectively. In the matrix notation

$$\boldsymbol{Y}_i = \boldsymbol{X}_{1i}\boldsymbol{\beta}_1 + \boldsymbol{Z}_{1i}\boldsymbol{b}_{1i} + \boldsymbol{\varepsilon}_i \ , \tag{3.1}$$

where in this notation $\boldsymbol{Y}_i$ is the longitudinal vector of response variable for the $i^{th}$ subject. $\boldsymbol{X}_{1i} = \big(\boldsymbol{x}_{1i}(s_{i1}), ..., \boldsymbol{x}_{1i}(s_{in_i})\big)'$ and $\boldsymbol{Z}_{1i} = \big(\boldsymbol{z}_{1i}(s_{i1}), ..., \boldsymbol{z}_{1i}(s_{in_i})\big)'$. We assume that $\boldsymbol{\varepsilon}_i \overset{\text{iid}}{\sim} SN_{n_i}\Big(-\sqrt{\frac{2}{\pi}}\boldsymbol{\delta}_e, \boldsymbol{\Psi}, \boldsymbol{\Delta}_e\Big)$ and $\boldsymbol{b}_{1i} \overset{\text{iid}}{\sim} N_{q_1}(0, \boldsymbol{D}_1)$. Note that these assumptions gives $E[\boldsymbol{\varepsilon}_i] = E[\boldsymbol{b}_{1i}] = \boldsymbol{0}$. Thus, this model considers the random effects $\boldsymbol{b}_{1i}$ to be symmetrically distributed, while the distribution of the within subject errors $\boldsymbol{\varepsilon}_i$ to be asymmetric with mean zero. To seek for identifiability (Arellano-Valle *et al.*, 2007), we assume $\boldsymbol{\Psi} = \sigma_e^2\boldsymbol{I}_{n_i}$, also, $\boldsymbol{\Delta}_e = \delta_e\boldsymbol{I}_{n_i}$.

In survival model, let $T_i^*$ be the true event time and $C_i$ be the censoring time. $T_i = \min(T_i^*, C_i)$ denotes the observed survival time for the $i^{th}$ individual, $i = 1, 2, ..., n$. Also, $\delta_i = I(T_i^* \leq C_i)$ is a censoring indicator, which is 0 for right-censored and 1 for complete observed individuals. Therefore, the observed data for the survival outcome consist of the pairs $\big\{(T_i, \delta_i), \ i = 1, 2, ..., n\big\}$.

For survival modeling a frailty model which is linked to the longitudinal model through some shared random effects is considered. The hazard function in

our proposed model is given by:

$$h\big(t_i|\boldsymbol{x}_{2i},\boldsymbol{z}_{2i},\boldsymbol{b}_{2i}\big) = h_0(t_i)\exp\big\{\boldsymbol{x}'_{2i}\boldsymbol{\beta}_2+\boldsymbol{z}'_{2i}\boldsymbol{b}_{2i}\big\}, \tag{3.2}$$

where $h_0(t_i)$ is the baseline hazard function. Thus, the density function of survival time for the $i^{th}$ individual can be written as:

$$h^{\delta_i}\big(t_i|\boldsymbol{x}_{2i},\boldsymbol{z}_{2i},\boldsymbol{b}_{2i}\big)\times\exp\Big\{-H_0(t_i)\exp\big\{\boldsymbol{x}'_{2i}\boldsymbol{\beta}_2+\boldsymbol{z}'_{2i}\boldsymbol{b}_{2i}\big\}\Big\},$$

where $H_0(t)=\int_0^t h_0(u)\,du$, $\boldsymbol{x}_2$ and $\boldsymbol{z}_2$ are $p_2$- and $q_2$-dimensional vectors of explanatory variables, respectively. $\boldsymbol{\beta}_2=(\beta_{21},...,\beta_{2p_2})'$ is a $p_2$-dimensional vector of time to event fixed effect parameters and $\boldsymbol{b}_{2i}=(b_{2i1},...,b_{2iq_2})'$ is a $q_2$-dimensional of random effects of time to event process where we assume, $\boldsymbol{b}_{2i}\overset{\text{iid}}{\sim}N_{q_2}(0,\boldsymbol{D}_2)$. In the next section, for getting sure of identifiability we shall impose $\boldsymbol{b}_{2i}$ to share some components with $\boldsymbol{b}_{1i}$.

The longitudinal outcome $\boldsymbol{y}_i$ can be partitioned into $\boldsymbol{y}_{i,\text{obs}}=\{y_i(s_{ij})\colon s_{ij}<T_i,\ j=1,2,...,n_i\}$, which contains all observed longitudinal measurements for the $i^{th}$ individual before the observed event time $T_i$, and $\boldsymbol{y}_{i,\text{mis}}=\{y_i(s_{ij})\colon s_{ij}\geq T_i,\ j=1,2,...,n_i\}$ which contains the longitudinal measurements that should have been taken until the end of the study and has to be considered as a vector containing missing values.

## 4. THE SKEW-NORMAL JOINT MODELING (SNJM) OF LONGITUDINAL AND SURVIVAL DATA

In our proposed joint modeling, we have considered a skew-normal distribution for error terms of longitudinal measurements. However, because of non-identifiability of some parameters, we have not considered skew-normal distribution assumption for random effects in these models. The skew-normal joint modeling (SNJM) of longitudinal and survival data, as an extension of the usual normal joint modeling, leads us to the following hierarchical model:

$$\begin{aligned}
\boldsymbol{Y}_i|\boldsymbol{b}_{1i},\boldsymbol{\beta}_1,\sigma_e^2,\delta_e &\overset{\text{ind.}}{\sim} SN_{n_i}\Big(\boldsymbol{X}_{1i}\boldsymbol{\beta}_1+\boldsymbol{Z}_{1i}\boldsymbol{b}_{1i}-\sqrt{\frac{2}{\pi}}\,\delta_e\mathbf{1}_{n_i},\,\sigma_e^2\boldsymbol{I}_{n_i},\,\delta_e\boldsymbol{I}_{n_i}\Big),\\
\boldsymbol{b}_{1i}|\boldsymbol{D}_1 &\overset{\text{ind.}}{\sim} N_{q_1}(\mathbf{0},\boldsymbol{D}_1),\\
h\big(t_i|\boldsymbol{x}_{2i},\boldsymbol{z}_{2i},\boldsymbol{b}_{2i}\big) &= h_0(t)\exp\big\{\boldsymbol{x}'_{2i}\boldsymbol{\beta}_2+\boldsymbol{z}'_{2i}\boldsymbol{b}_{2i}\big\},\\
\boldsymbol{b}_{2i}|\boldsymbol{D}_2 &\overset{\text{ind.}}{\sim} N_{q_2}(\mathbf{0},\boldsymbol{D}_2),
\end{aligned} \tag{4.1}$$

where some components of random effects are shared between two models. The stochastic representation of the skew-normal distribution can be used for simplifying Markov Chain Monte Carlo (MCMC) approach in Bayesian specification.

Therefore, the first line of equation (4.1) can be written as:

$$\begin{aligned} \boldsymbol{Y}_i|\boldsymbol{b}_{1i},\boldsymbol{\beta}_1,\sigma_e^2,\boldsymbol{U}_i = \boldsymbol{u}_i &\overset{\text{ind.}}{\sim} N_{n_i}\Big(\boldsymbol{X}_{1i}\boldsymbol{\beta}_1+\boldsymbol{Z}_{1i}\boldsymbol{b}_{1i}+\delta_e\boldsymbol{u}_i-\sqrt{\frac{2}{\pi}}\,\delta_e\mathbf{1}_{n_i},\,\sigma_e^2\boldsymbol{I}_{n_i}\Big)\,,\\ \boldsymbol{U}_i &\overset{\text{ind.}}{\sim} N_{n_i}(\mathbf{0},\boldsymbol{I}_{n_i})\,I(\boldsymbol{u}_i>0)\,, \end{aligned}$$

where $\boldsymbol{u}_i$ is the observed value of $\boldsymbol{U}_i$. Some components of longitudinal measurements may be missing due to dropout. We consider a non-ignorable missingness mechanism for them. In order to complete the Bayesian specification, prior distributions for all unknown parameters should be defined. The vector of unknown parameter is $\boldsymbol{\theta}=(\boldsymbol{\beta}_1',\boldsymbol{\beta}_2',\sigma_e^2,\delta_e,\boldsymbol{D}_1,\boldsymbol{D}_2)$. To ensure to have proper posteriors in the model we consider proper but diffuse conditionally conjugate priors (Hobert and Casella, 1996). We assume that components of $\boldsymbol{\theta}$ are mutually independent and the prior distributions are given by

$$\begin{aligned} &\boldsymbol{\beta}_1\sim N_{p_1}(\boldsymbol{\beta}_{01},\boldsymbol{\Sigma}_{01})\,, && \boldsymbol{\beta}_2\sim N_{p_2}(\boldsymbol{\beta}_{02},\boldsymbol{\Sigma}_{02})\,,\\ &\boldsymbol{D}_1\sim IW_{q_1}(\boldsymbol{\eta}_{01},\boldsymbol{\psi}_{01})\,, && \boldsymbol{D}_2\sim IW_{q_2}(\boldsymbol{\eta}_{02},\boldsymbol{\psi}_{02})\,,\\ &\sigma_e^2\sim I\Gamma(\alpha_0,\tau_0)\,, && \delta_e\sim N(\mu_{\delta_e},\sigma_{\delta_e}^2)\,. \end{aligned} \tag{4.2}$$

The hyperparameters of these priors are selected such that they lead to the low-informative prior distributions. As all of these priors are proper but, low-informative in view of their variances.

For Bayesian implementation, one may use Gibbs sampling and Metropolis–Hastings algorithm via WinBUGS package.

### 4.1. Models comparison

For models comparison, we have used some famous criteria which are Deviance Information Criterion (DIC), Expected Akaike Information Criterion (EAIC), Expected Bayesian Information Criterion (EBIC, Carlin and Louis, 2000; Brooks, 2002) and Log Pseudo Marginal Likelihood (LPML).

Let $\boldsymbol{\Theta}$ and $\boldsymbol{Z}=(z_1,...,z_N)'$ be the entire model parameters and data, respectively. Define: $D(\boldsymbol{\Theta})=-2\ln f(\boldsymbol{z}|\boldsymbol{\Theta})=-2\sum_{i=1}^{N}\ln f(z_i|\boldsymbol{\Theta})$, where $f(z_i|\boldsymbol{\Theta})$ is marginal distribution of $z_i$, then $E\left[D(\boldsymbol{\Theta})\right]$ is a measure of fit and can be approximated by using the MCMC output in a Monte Carlo integration. This index is given by $\bar{D}=\frac{1}{K}\sum_{k=1}^{K}D(\boldsymbol{\Theta}^{(k)})$. Where $\boldsymbol{\Theta}^{(k)}$ is the $k^{th}$ iteration of MCMC chain of the model and $K$ is the number of iterations.

Therefore the Bayesian criteria are given by $\widehat{DIC}=\bar{D}+\hat{p}_D$, $\widehat{EAIC}=\bar{D}+2p$ and $\widehat{EBIC}=\bar{D}+p\ln(N)$, where $p$ is the number of parameters and $N$ is the

total number of observations. The smaller DIC, EAIC and EBIC, the better fit of the model.

Another popular criterion, which is usually used for model comparison in Bayesian context is Conditional Predictive Ordinate (CPO) statistic. Let $\boldsymbol{Z}^{(-i)}$, $i = 1, 2, ..., N$, denote the data set without its $i^{th}$ individual, and let $\pi(\boldsymbol{\theta}|\boldsymbol{Z}^{(-i)})$ denote posterior distribution of $\boldsymbol{\theta}$ given $\boldsymbol{Z}^{(-i)}$, then $CPO_i = \int_{\boldsymbol{\Theta}} f(z_i|\boldsymbol{\theta})\pi(\boldsymbol{\theta}|\boldsymbol{z}^{(-i)})\, d\boldsymbol{\theta}$. Gelfand and Dey (1994) show that $CPO_i$ can be estimated by

$$CPO_i = \left( \frac{1}{K} \sum_{k=1}^{K} \frac{1}{f\left(z_i|\boldsymbol{\theta}^{(k)}, \boldsymbol{z}\right)} \right)^{-1} .$$

For collecting information of $CPO_i$s, the LPML statistic is defined by $LPML = \sum_{i=1}^{N} \log(CPO_i)$. In this concept, unlike that of DIC, EAIC or EBIC, the larger value of LPML criterion indicates a better fitted model.

### 4.2. Convergence diagnostics

Gelman and Rubin (1992) have suggested a diagnostic test for assessing convergence. Their method recommends that two or more parallels (denoted by $m$) chains be generated, each with different starting values. For assessing convergence of individual model parameters the potential scale reduction factor (PSRF) may be used. The PSRF is calculated by $PSRF = \sqrt{\frac{n-1}{n} + \frac{m+1}{nm}\frac{B}{W}}$, where $B/n$ is the between-chain variance $\left[\frac{\mathrm{B}}{n} = \frac{1}{m-1}\sum_{j=1}^{m}\left(\theta_j - \bar{\theta}\right)^2\right]$ and $W$ is the within-chain variance $\left[\mathrm{W} = \frac{1}{m(n-1)}\sum_{j=1}^{m}\sum_{i=1}^{n}\left(\theta_{ij} - \bar{\theta}_j\right)^2\right]$. As chains converge to a common target distribution, the between-chain variability should become small relative to the within-chain variability and consequently $PSRF$ should be close to 1. Conversely, PSRF value larger than 1 indicates non-convergence.

## 5. SIMULATION STUDY

To investigate the performance of the proposed model, we conducted a simulation study. In this simulation study, we generate 500 samples with sample size $n = 100$, moderate sample size, and $n = 500$, large sample size. We have considered the following joint modeling:

$$y_{ij} = \beta_{11} + \beta_{12}s_{ij} + \beta_{13}x_i + b_{1i} + b_{2i}s_{ij} + \varepsilon_{ij} , \qquad i = 1, 2, ..., n , \quad j = 1, 2..., 5 .$$

In this model $s_{ij} = j$, $x_i \sim Ber(0.2)$, $\beta_{11} = 10$, $\beta_{12} = -3$, $\beta_{13} = -2$, $\varepsilon_{ij} \sim SN\Big(-\sqrt{\frac{2}{\pi}}\,\delta_e, \sigma_e^2, \delta_e\Big)$, where $\sigma_e = 1$ and $\delta_e = 3$. Also, we have used a Cox proportional hazard model in a frailty structure with a Weibull baseline hazard as follows:

$$h(t) \;=\; h_0(t)\exp\big\{\beta_{21} + \beta_{22}x_i + \rho_1 b_{1i} + \rho_2 b_{2i}\big\}\,.$$

In this model, $\beta_{22} = -2$, $\rho_1 = 1$ and $\rho_2 = 2$. We have considered three rates of random dropout, 10%, 30% and 50%, which are generated by using different values for $\beta_{21}$. Therefore, we have a non-ignorable mechanism in the model, such that when $s_{ij} > T_i$ then $i^{th}$ individual dropouts from the study. In this simulation study $\beta_{21} = 3$ leads to 10% rate of non-random dropout, $\beta_{21} = -1$ leads to 30% and $\beta_{21} = -2$ leads to 50% rate of non-random dropout of longitudinal outcomes. Also, $\boldsymbol{b}_i = (b_{1i}, b_{2i}) \sim N_2(0, \boldsymbol{D})$, where $\boldsymbol{D}$ is considered to be a $2\times2$ matrix, where $d_{11} = d_{22} = 1$ and $d_{21} = d_{12} = 0.5$. $d_{11}$, $d_{12}$ and $d_{22}$ are distinct elements of the matrix $\boldsymbol{D}$. Let $\boldsymbol{\beta}_1 = (\beta_{11}, \beta_{12}, \beta_{13})'$ and $\boldsymbol{\beta}_2 = (\beta_{21}, \beta_{22})'$, we have used the following low-informative priors for unknown parameters.

$$\begin{aligned}
&\boldsymbol{\beta}_1 \sim N_3(0, 10^3 I_3)\,, && \boldsymbol{\beta}_2 \sim N_2(0, 10^3 I_2)\,,\\
&\sigma_e \sim I\Gamma(0.01, 0.01)\,, && \delta_e \sim N(0, 100)\,,\\
&\rho_k \sim N_2(0, 10^2 I),\ \ k = 1, 2\,, && \boldsymbol{D} \sim IW_2(100 I_2, 2)\,.
\end{aligned} \tag{5.1}$$

We have used "R2WinBUGS" package for implementation of this simulation study. We have implemented 10,000 iterations and have used the last 5000 iterations to obtain some summary of posterior. We have analysed these simulated data set under two distributional assumptions for within subject error: normal distributional assumption and skew-normal distributional assumption. The results of this simulation study are presented in Tables 1 to 3 for the rate of missingness 10%, 30% and 50%, respectively. The relative bias and mean square error of parameter $\boldsymbol{\theta}$ are defined as

$$Rel.Bias(\theta) = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{\hat{\theta}_i}{\theta} - 1\right), \qquad MSE(\theta) = \frac{1}{N}\sum_{i=1}^{N}\big(\hat{\theta}_i - \theta\big)^2,$$

where $\hat{\theta}_i$ is the estimate of $\theta$ for the $i^{th}$ sample and $N = 500$. These tables show that when real data have skew-normal distribution, some parameters of joint modeling under normal assumption are estimated with some biases. These parameters are variance of error term of longitudinal model, variance of random effects and coefficients of random effects of survival model. For comparison of the performance of two models results of relative bias and mean square error of estimators are considered. Based on results given in Tables 1–3, we can conclude that skew-normal model leads to better inference in general. Also, these tables show that increasing of sample size in skew-normal scenario is an effective measure in decreasing standard errors, relative bias and MSE of the parameters.

## 6. HIV DATA SET

As an illustrative example of our Bayesian joint modeling, we use a longitudinal study on 467 HIV patients. Data were collected by Goldman *et al.* (1996). HIV infection results in a progressive destruction of immune function, which may be indicated by a decrease of CD4 (Stevens *et al.*, 2006). A count of CD4 cells of a person gives a general measure of the health of him/her immune system, and is a good measurement of immunosuppression. A normal CD4 cell count is more than 500 cells per cubic millimeter (mm3) of blood. If one has a CD4 count of fewer than 300, one will be diagnosed as having AIDS, therefore, CD4 count measurement is an important index which provides a way of gauging the progression from HIV to acquired immune deficiency syndrome (AIDS) for prognostic purposes. Thus, in this study, the CD4 count measurements over time are chosen as response variable.

This study is done for comparing the efficacy and safety of two alternative antiretroviral drugs, namely didanosine (ddI) and zalcitabine (ddC). The patients met another entry conditions which is AIDS diagnosis or two CD4 counts of 300 or fewer, also they randomly assigned to receive either ddI or ddC, and CD4 cell counts were recorded at study entry and again at the 2, 6, 12, and 18 months. Another variable which is recorded in this study is time to death.

Before this Guo and Carlin (2004), Rizopoulos (2010) and some other authors had suggested that, for analysing this data set, a square root transformation for CD4 counts be used instead of Gaussian model. Figure 1 shows histogram and q-q plot of $\sqrt{CD4}$ which shows that there are right skewness $\sqrt{CD4}$ even after root transformation. We have used the same transformation in our analysis, but under skew-normal distribution assumption for the error term. Figure 2 presents the subject-specific profile for fifty randomly selected individuals given each drug. Panels of this figure show a sharply increasing degree of missing data over time due to death, dropout, and missed clinic visits. In this figure the profiles of those individuals who remain and those of individuals who do not remain are indicated using gray and black colors, respectively. This figure underlines that those who do not remain had smaller $\sqrt{CD4}$ than others. Figure 3 presents Kaplan–Meier survival curve estimates for both treatment groups. The plot suggests longer survival times in the ddC group compared to the ddI group, from month 6 onwards.

We have used a skew-normal JM of longitudinal and time to event data for analysing the data set. The linear mixed effect model with random intercept and slope is:

$$\begin{aligned} y_{ij} \;=\; & \beta_{11} + \beta_{12}\, t_{ij} + \beta_{13}\, t_{ij}\, Drug_i + \beta_{14}\, Gender_i \\ & +\; \beta_{15}\, PrevOI_i + \beta_{16}\, Stratum_i + b_{1i} + b_{2i}\, t_{ij} + \sigma_e\, \varepsilon_{ij} \;. \end{aligned} \tag{6.1}$$

For the time to event process, we have used a Cox proportional hazard model, the hazard function for this model is given by

$$h(t_i) = h_0(t_i)\exp\Big\{\beta_{21} + \beta_{22}\,Drug_i + \beta_{23}\,Gender_i + \beta_{24}\,PrevOI_i + \beta_{25}\,Stratum_i + \rho_1 b_{1i} + \rho_2 b_{2i}\Big\}. \tag{6.2}$$

In models (6.1) and (6.2) the vector of random effects $\boldsymbol{b}_i = (b_{1i}, b_{2i})'$ is shared between two models. Also, we consider normal and skew-normal distribution assumptions on the longitudinal mixed model. Random effects are assumed to have a bivariate normal distribution, that is, $\boldsymbol{b}_i \sim N_2(\boldsymbol{0},\boldsymbol{D})$ and $\varepsilon_{ij} \sim SN\Big(-\sqrt{\frac{2}{\pi}}\delta_e, \sigma_e^2, \delta_e\Big)$. In this model, $y_{ij}$ is the squared root of the $j^{th}$ CD4 count measurement on the $i^{th}$ individual in the trial, $j = 1, 2, ..., 5$ and $i = 1, 2, ..., 467$. $Gender_i$ is a gender indicator ($0 =$ female, $1 =$ male), also other three explanatory variables are $Drug_i$ ($0 =$ ddC, $1 =$ ddI), $PrevOI_i$, previous opportunistic infection, ($1 =$ AIDS diagnosis, $0 =$ no AIDS diagnosis), and $Stratum_i$ ($1 =$ AZT failure, $0 =$ AZT intolerance).

In Bayesian MCMC implementation, we ran two parallels MCMC chains with different starting values for $100,000$ iterations each. Then, we discarded the first $20,000$ iterations as pre-convergence burn-in and retained $80,000$ for the posterior analysis. Let $\boldsymbol{\beta}_k = (\beta_{k1}, ..., \beta_{kp_k})'$ where $k = 1, 2,\ p_1 = 6,\ p_2 = 5$. In all models, we consider $\boldsymbol{\beta}_k \sim N_{p_k}(0, 10000 I_{p_k})$, $\sigma_e^2 \sim I\Gamma(0.1, 0.1)$, $\rho_k \sim N(0, 100)$, $\boldsymbol{D} \sim IW_2(100 I_2, 2)$ and $\delta_e \sim N(0, 100)$. For the piecewise baseline hazard function [$h_l,\ l = 1, 2, 3, 4$ (the number of piecewise baseline $= 4$)] the $gamma(1, 1)$ prior distribution is considered for each piece ($h_i$, $i = 1, 2, 3, 4$). Hyperparameters are chosen such that the priors of the parameters tend to be weakly informative. We have considered joint modeling of equations (7)–(8) under two different distribution assumptions.

After checking Gelman–Rubin diagnosis test for convergence, Bayesian parameter estimates including posterior mean, standard deviation and 95% highest posterior density of all parameters are given in Table 4. According to DIC, EAIC, EBIC and LPML criteria the skew-normal model has a better fit to these data. This table shows that time and previous opportunistic infection are two significant covariates in longitudinal model, also skewness parameter of error term is significant, where the more time and previous opportunistic infection, the less CD4 count measurements. In survival model $\rho_1$ and $\rho_2$ are significant which shows that two models are dependent. Also, Table 4 shows that skewness parameter is significant and ignoring this parameter and using normal model leads to overestimating of variance of the error in longitudinal model.

An important criterion for finding the influential observations is Kullback–Leibler divergence criterion between $\pi(\boldsymbol{\theta}|\boldsymbol{Z})$ and $\pi(\boldsymbol{\theta}|\boldsymbol{Z}^{(-i)})$, $i = 1, 2, ..., n$, where $\boldsymbol{Z}$ and $\boldsymbol{Z}^{(-i)}$ are all data and the data set without its $i^{th}$ individual, respectively.

It is defined by

$$K_i = \int \pi(\boldsymbol{\theta}|\boldsymbol{y},\boldsymbol{t}) \log\left(\frac{\pi(\boldsymbol{\theta}|\boldsymbol{y},\boldsymbol{t})}{\pi\big(\boldsymbol{\theta}|\boldsymbol{y}^{(-i)},\boldsymbol{t}^{(-i)}\big)}\right) d\boldsymbol{\theta} \ .$$

This can be approximated by: (Christensen *et al.*, 2011; page 341)

$$K_i = \log\left(\frac{1}{m}\sum_{j=1}^{m}\frac{1}{L\big(\boldsymbol{\theta}_j|\boldsymbol{y}_i,t_i\big)}\right) - \frac{1}{m}\sum_{j=1}^{m}\log\left(\frac{1}{L\big(\boldsymbol{\theta}_j|\boldsymbol{y}_i,t_i\big)}\right).$$

An observation with large $K_i$ is considered as an influential observation. Figure 4 shows Kullback–Leibler divergence for both skew-normal and normal models. This figure shows that the skew-normal model detects some individuals as influential observations, but normal model does not detect any. Individuals 131 and 353 have large values of the response in the second observed time in comparison with the largest value in this time. These individuals have observed survival times 12.23 and 12.53, respectively. There is so much increase in CD4 count measurements at $3^{th}$ to $4^{th}$ observed time for individuals 188 and 245. Other individuals detected in Figure 3 are 417, 171 and 319. Except observation 417 who is died at time 10.60, the other individuals dropped out from the study at times 17.33 and 15.97, respectively, where the length of the period of the study is equals to 18. The longitudinal measurements for individuals 417 and 171 are close to the lowest value of the CD4 measurements at each time of the study.

Also, a sensitivity analysis is performed to see the modification of posterior distribution with respect to changes in the hyperparameters of prior distributions of $\sigma_e^2$. For this purpose, we assume $\sigma_e^2 = \frac{1}{\tau_e^2}$, where $\tau_e^2 \sim \Gamma(\epsilon,\epsilon)$ (see Gelman, 2006) and $\mathrm{Var}(\tau_e^2) = \frac{1}{\epsilon} = 10^k$, $k = -3,-2,...,2,3$. Sensitivity of the posterior mean of all parameters for different values of $k$ is investigated. This shows that our inferences, containing the results of posterior means, standard deviations and DIC value, are not sensitive to the change of value of $k$ after $k = 1$ (the results are not given here to save space).

## 7. CONCLUSION

In this paper, we have used a multivariate skew-normal distribution family, which includes normal distribution, for analysing skew longitudinal responses in joint modeling of longitudinal and survival times. We have used Bayesian approach and WinBUGS software for implementing the proposed model. We have performed some simulation studies to investigate the performance of the proposed method. Also, the proposed method is used for analysing a real HIV data set. Our analysis shows that these data set are severely skew, and the skew-normal model has a better performance than normal model based on the

DIC, EAIC, EBIC and LPML criteria. The program codes for analysing the data set are available under request from the authors. If a data set includes outliers and skew longitudinal responses a joint model with assumption of skew-normal/independent distribution for responses may be defined to analyse the data.

## REFERENCES


[1] Arellano-Valle, R.B.; Bolfaine, H. and Lachos, V.H. (2005a). Skew-normal linear mixed models, *Journal of Data Science*, **3**, 415–438.

[2] Arellano-Valle, R.B. and Genton, M.G. (2005). On fundamental skew distributions, *Journal of Multivariate Analysis*, **96**, 93–116.

[3] Arellano-Valle, R.B.; Ozan, S.; Bolfarine, H. and Lachos, V.H. (2005b). Skew-normal measurement error models, *Journal of Multivariate Analysis*, **96**, 265–281.

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

[5] Azzalini, A. (1985). A class of distribution which includes the normal ones, *Scandinavian Journal of Statistics*, **12**, 171–178.

[6] Azzalini, A. (1986). Further results on a class of distributions which includes the normal ones, *Statistica XLVI*, 199–208.

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

[8] Azzalini, A. and Dalla-Valle, A. (1996). The multivariate skew-normal distribution, *Biometrika*, **83**, 715-726.

[9] Baghfalaki, T. and Ganjali, M. (2011). An EM estimation approach for analyzing bivariate skew-normal data with non-monotone missing values, *Communications in Statistics – Theory and Methods*, **40**(9), 1671–1686.

[10] Baghfalaki, T. and Ganjali, M. (2012). An ECM Estimation Approach for Analyzing Multivariate Skew-Normal Data with Dropout, *Communications in Statistics – Simulation and Computation*, **41**, 1970–1988.

[11] Bandyopadhyay, D.; Lachos, V.H. and Abanto-Valle, C.A. (2010). Bayesian inference for bivariate skew-normal/independent linear mixed models with application to periodontal disease, *Statistics in Medicine*, **29**, 2643–2655.

[12] Branco, M.D. and Dey, D.K. (2001). A general class of multivariate skew-elliptical distributions, *Journal of Multivariate Analysis*, **79**, 99–113.

[13] Brooks, S.P. (2002). Discussion on the paper by Spiegelhalter Best Carlin and vandeLinde, 2002, *Journal of the Royal Statistical Society Series B*, **64**(3), 616–618.

[14] Cancho, V.G.; Lachos, V.H. and Ortega, E.M. (2010). A nonlinear regression model with skew-normal errors, *Statistical Papers*, **51**, 547–558.

[15] Carlin, P. and Louis, T.A. (2000). *Bayes and Empirical Bayes Methods for Data Analysis*, Second Edition, Chapman & Hall, CRC Press, New York.

[16] Christensen, R.; Johnson, W.; Branscum, A. and Hanson, T.E. (2011). *Bayesian Ideas and Data Analysis, An Introduction for Scientists and Statisticians*, Chapman & Hall / CRC.

[17] DeGruttola, V. and Tu, X.M. (1994). Modeling Progression of CD4-Lymphocyte Count and Its Relationship to Survival Time, *Biometrics*, **50**, 1003–1014.

[18] Diggle, P.J.; Sousa, I. and Chetwynd, A.G. (2008). Joint modelling of repeated measurements and time-to-event outcomes: the fourth Armitage lecture, *Statistics in Medicine*, **27**, 2981–2998.

[19] Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models (Comment on article by Browne and Draper), *Bayesian Analysis*, **1**, 515–534.

[20] Gelman, A. and Rubin, D.B. (1992). Inference from Iterative Simulation Using Multiple Sequences, *Statistical Science*, **7**, 457–511.

[21] Goldman, A.I.; Carlin, B.P.; Crane, L.R.; Launer, C.; Korvick, J.A.; Deyton, L. and Abrams, D.I. (1996). Response of CD4+ and Clinical Consequences to Treatment Using ddI or ddC in Patients with Advanced HIV Infection, *Journal of Acquired Immune Deficiency Syndromes and Human Retrovirology*, **11**, 161–169.

[22] Ghosh, P. and Branco, D.B. (2007). Bivariate random effect model using skew-normal distribution with application to HIV-RNA, *Statistics in Medicine*, **26**(6), 1255–1267.

[23] Guo, X. and Carlin B.P. (2004). Separate and Joint Modeling of Longitudinal and Event Time Data Using Standard Computer Packages, *The American Statistician*, **58**, 16–24.

[24] Hashemi, R.; Jacqmin-Gadda, H. and Commenges, D. (2003). A latent process model for joint modeling of events and markers, *Lifetime Data Analysis*, **8**(4), 335–348.

[25] Henderson, R.; Diggle, P.J. and Dobson, A. (2000). Joint Modelling of Longitudinal Measurements and Event Time Data, *Biostatistics*, **1**, 465–480.

[26] Henze, N.A. (1986). A probabilistic representation of the skew-normal distribution, *Scandinavian Journal of Statistics*, **13**, 271–275.

[27] Hobert, J. and Casella, G. (1996). The Effect of Improper Priors on Gibbs Sampling in Hierarchical Linear Mixed Models, *Journal of the American Statistical Association*, **91**, 1461–1473.

[28] Huang, Y.; Chen, R. and Dagne, G. (2011a). Simultaneous Bayesian Inference for Linear, Nonlinear and Semiparametric Mixed-Effects Models with Skew-Normality and Measurement Errors in Covariates, *The International Journal of Biostatistics*, **7**(1), Article 8.

[29] Huang, Y. and Dagne, G. (2010). Skew-normal Bayesian nonlinear mixed-effects models with application to AIDS studies, *Statistics in Medicine*, **29**(23), 2384–2398.

[30] Huang, Y. and Dagne, G. (2011). A Bayesian Approach to Joint Mixed-Effects Models with a Skew-Normal Distribution and Measurement Errors in Covariates, *Biometrics*, **67**, 260–269.

[31] Huang, Y.; Dagne, G. and Wu, L. (2011b). Bayesian inference on joint models of HIV dynamics for time-to-event and longitudinal data with skewness and covariate measurement errors, *Statistics in Medicine*, **30**(24), 2930–2946.

[32] Lachos, V.H.; Bolfarine, H.; Arellano-Valle, R.B. and Montenegro, L.C. (2007). Likelihood based inference for multivariate skew-normal regression models, *Communications in Statistics – Theory and Methods*, **36**, 1769–1786.

[33] Lachos, V.H.; Bandyopadhyay, D. and Dey, D.K. (2011). Linear and Nonlinear Mixed-Effects Models for Censored HIV Viral Loads Using Skew-Normal/Independent Distributions, *Biometrics*, **67**, 1594–1604.

[34] Lachos, V.H.; Ghosh, P. and Arellano-Valle, R.B. (2010). Likelihood Based Inference for Skew-Normal/Independent Linear Mixed Model, *Statistica Sinica*, **20**, 303–322.

[35] Lin, I.; Ho, J. and Chen, L. (2009). Analysis of multivariate skew-normal models with incomplete data, *Journal of Multivariate Analysis*, **100**, 2337–2351.

[36] Lin, T. and Lee, J. (2008). Estimation and prediction in linear mixed models with skew-normal random effects for longitudinal data, *Statistics in Medicine*, **27**, 1490–1507.

[37] Rizopoulos, D. (2010). JM: An R Package for the Joint Modelling of Longitudinal and Time-to-Event Data, *Journal of Statistical Software*, **35**(9).

[38] Sahu, S.K.; Dey, D.K. and Branco, M. (2003). A new class of multivariate skew distributions with applications to Bayesian regression models, *Canadian Journal of Statistics*, **31**(2), 129–150.

[39] Sousa, I. (2011). A review on joint modelling of longitudinal measurements and time-to-event, *REVSTAT – Statistical Journal*, **9**(1), 57–81.

[40] Spiegelhalter, D.J.; Best, N.G.; Carlin, B.P. and Lindevan der, A. (2002). Bayesian measures of model complexity and fit, *Journal of the Royal Statistical Society Series B*, **64**, 583–616.

[41] Stevens, R.A.; Lempicki, R.A.; Natarajan, V.; Higgins, J.; Adelsberger, J.W. and Metcalf, J.A. (2006). *General immunologic evaluation of patients with immunodeffciency virus infection.* In "Manual of Molecular and Clinical Laboratory Immunology" (B. Detrick, R.G. Hamilton, J.D. Folds, *et al.*, Eds.), 7th edition, 848–861, Washington, DC, ASM Press.

[42] Tseng, Y.; Hsieh, F. and Wang, J. (2005). Joint modelling of accelerated failure time and longitudinal data, *Biometrika*, **92**, 587–603.

[43] Tsiatis, A.A. and Davidian, M. (2001). A semiparametric estimator for the proportional hazards model with longitudinal covariates measured with error, *Biometrika*, **88**, 447–458.

[44] Tsiatis, A.A.; DeGruttola, V. and Wulfsohn, M.S. (1995). Modeling the relationship of survival to longitudinal data measured with error. Applications to survival and CD4 counts in patients with AIDS, *Journal of the American Statistical Association*, **90**(429), 27–37.

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

[46] Wulfsohn, M.S. and Tsiatis, A.A. (1997). A Joint Model for Survival and Longitudinal Data Measured with Error, *Biometrics*, **53**, 330–339.

[47] Yu, M.; Law, N.J.; Taylor, J.M.G. and Sandler, H.M. (2004). Joint longitudinal survival cure models and their application to prostate cancer, *Statistica Sinica*, **14**, 835–862.

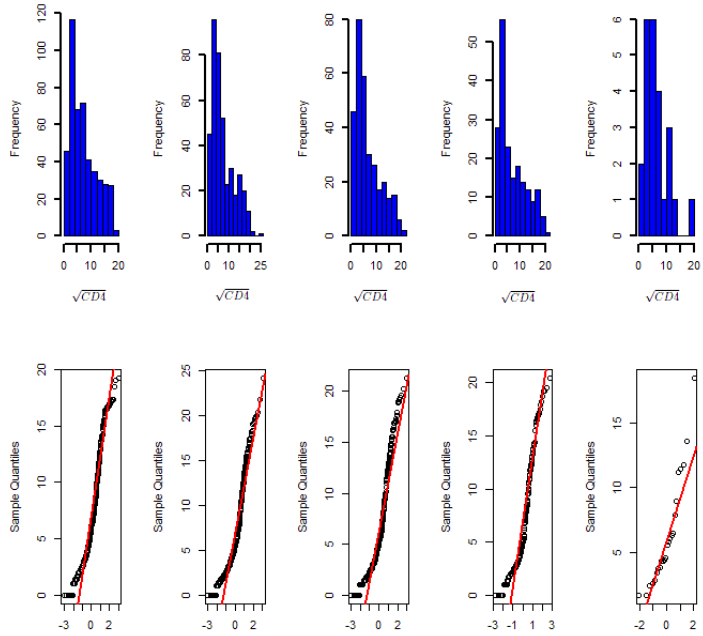


**Figure 1**: Histogram and q-q plot of $\sqrt{CD4}$ in HIV data set.

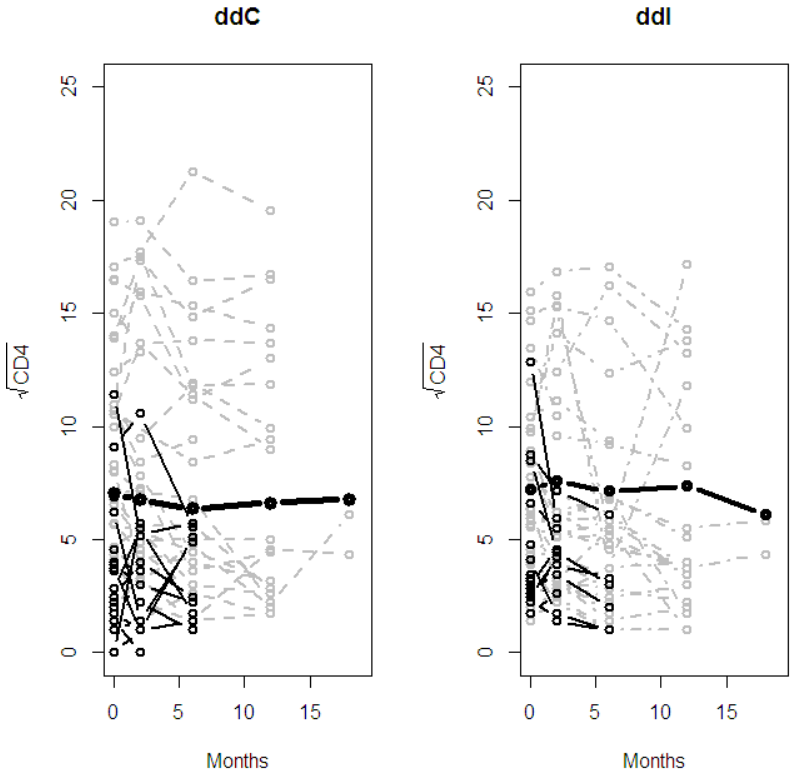


**Figure 2**: Profiles of $\sqrt{CD4}$ measurements over time for all individuals from each drug, bold black lines are mean profile for all observed individuals on each drug, gray color indicates those individuals who remain and black color represents those who do not remain.

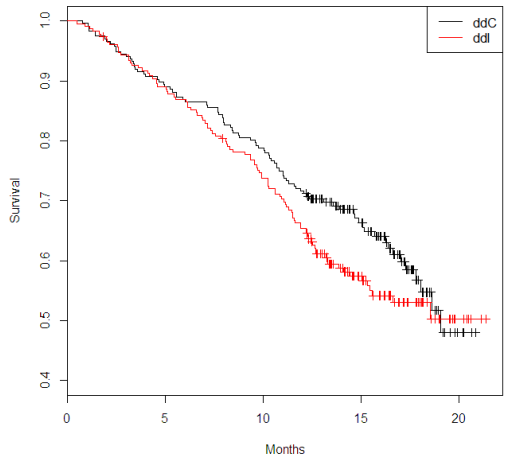


**Figure 3**: Kaplan–Meier estimates of the probability of survival for individuals on each drug.

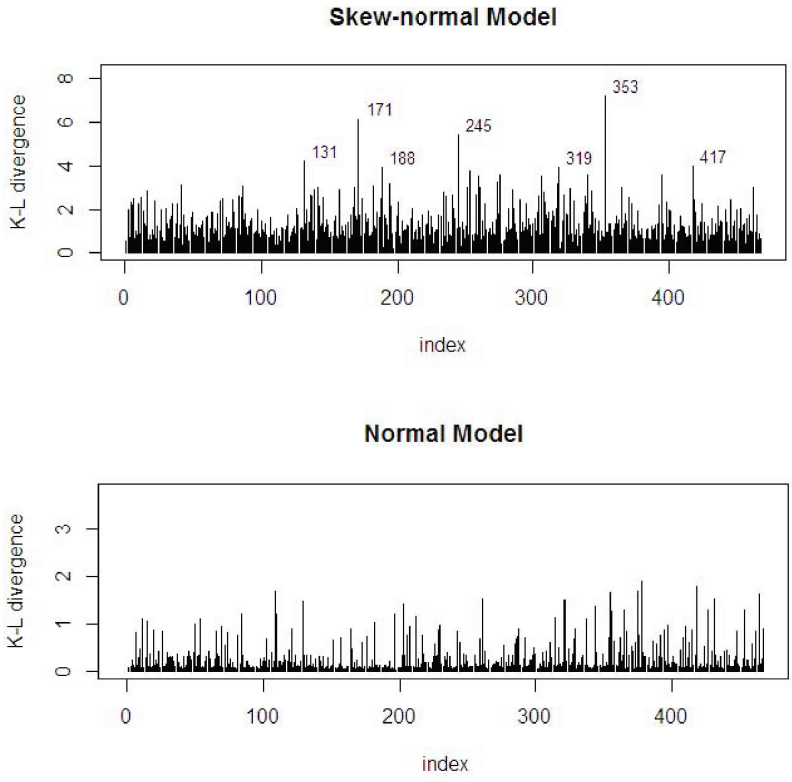


**Figure 4**: Kullback–Leibler divergence for the skew-normal and normal models.

**Table 1**: Result of simulation study for 500 samples under two distribution assumptions with 10% rate of random dropout. (Est.: posterior mean, S.E.: standard error, Rel. Bias: relative bias and MSE: mean square error).

| Model | | Normal model | | | | | | Skew-normal model | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Sample size | | $n = 100$ | | | $n = 500$ | | | $n = 100$ | | | $n = 500$ | | |
| parameters | real | Est.(S.E.) | Rel. Bias | MSE | Est.(S.E.) | Rel. Bias | MSE | Est.(S.E.) | Rel. Bias | MSE | Est.(S.E.) | Rel. Bias | MSE |
| $\beta_{11}$ | 10.000 | 9.966(0.292) | -0.003 | 0.086 | 9.965(0.091) | -0.003 | 0.009 | 10.027(0.206) | 0.002 | 0.042 | 10.003(0.122) | 0.0003 | 0.015 |
| $\beta_{12}$ | -3.000 | -3.013(0.189) | 0.004 | 0.035 | -2.995(0.047) | -0.001 | 0.002 | -3.010(0.129) | 0.003 | 0.016 | -2.989(0.072) | -0.003 | 0.005 |
| $\beta_{13}$ | -2.000 | -1.939(0.541) | -0.030 | 0.294 | -1.960(0.130) | -0.019 | 0.017 | -1.995(0.366) | -0.002 | 0.133 | -1.988(0.187) | -0.005 | 0.035 |
| $\beta_{21}$ | 3.000 | 2.928(0.526) | -0.023 | 0.281 | 2.977(0.151) | -0.007 | 0.022 | 2.964(0.356) | -0.012 | 0.127 | 3.012(0.196) | 0.004 | 0.038 |
| $\beta_{22}$ | -2.000 | -1.904(0.455) | -0.047 | 0.215 | -1.938(0.127) | -0.031 | 0.019 | -2.008(0.357) | 0.004 | 0.126 | -1.986(0.192) | -0.006 | 0.037 |
| $d_{11}$ | 1.000 | 5.055(0.462) | 4.055 | 16.659 | 5.961(0.143) | 3.961 | 20.943 | 1.487(0.287) | 0.487 | 0.319 | 1.236(0.172) | 0.236 | 0.085 |
| $d_{12}$ | 0.500 | 0.241(0.228) | -0.517 | 0.119 | 0.392(0.074) | -0.215 | 0.016 | 0.421(0.146) | -0.158 | 0.027 | 0.463(0.092) | -0.072 | 0.009 |
| $d_{22}$ | 1.000 | 3.212(0.240) | 2.212 | 4.952 | 3.254(0.071) | 3.254 | 6.069 | 1.141(0.152) | 0.141 | 0.043 | 1.054(0.090) | 0.054 | 0.011 |
| $\sigma_e^2$ | 1.000 | 4.146(0.521) | 3.146 | 10.168 | 4.067(0.154) | 3.067 | 9.434 | 0.919(0.525) | -0.080 | 0.280 | 0.906(0.243) | -0.093 | 0.067 |
| $\delta_e$ | 3.000 | - | - | - | - | - | - | 2.972(0.386) | -0.009 | 0.149 | 3.020(0.156) | 0.006 | 0.024 |
| $\rho_1$ | 1.000 | 0.521(0.166) | -0.478 | 0.256 | 0.684(0.070) | -0.315 | 0.104 | 0.793(0.187) | -0.206 | 0.077 | 0.885(0.125) | -0.114 | 0.028 |
| $\rho_2$ | 2.000 | 2.237(0.201) | 0.118 | 0.096 | 2.193(0.063) | 0.096 | 0.041 | 2.146(0.179) | 0.073 | 0.053 | 2.076(0.102) | 0.038 | 0.016 |

*$d_{11}$, $d_{12}$ and $d_{22}$ are distinct elements of the matrix $\boldsymbol{D}$.*

**Table 2**: Result of simulation study for 500 samples under two distribution assumptions with 30% rate of random dropout. (Est.: posterior mean, S.E.: standard error, Rel. Bias: relative bias and MSE: mean square error).

| Model | | Normal model | | | | | | Skew-normal model | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Sample size | | $n = 100$ | | | $n = 500$ | | | $n = 100$ | | | $n = 500$ | | |
| parameters | real | Est.(S.E.) | Rel. Bias | MSE | Est.(S.E.) | Rel. Bias | MSE | Est.(S.E.) | Rel. Bias | MSE | Est.(S.E.) | Rel. Bias | MSE |
| $\beta_{11}$ | 10.000 | 10.181(0.276) | 0.018 | 0.104 | 10.046(0.247) | 0.004 | 0.062 | 10.009(0.229) | 0.001 | 0.052 | 10.008(0.132) | 0.0008 | 0.017 |
| $\beta_{12}$ | -3.000 | -3.156(0.259) | 0.052 | 0.087 | -3.088(0.161) | 0.029 | 0.033 | -3.013(0.158) | 0.004 | 0.025 | -3.010(0.095) | 0.003 | 0.009 |
| $\beta_{13}$ | -2.000 | -1.905(0.849) | -0.047 | 0.686 | -1.866(0.407) | -0.066 | 0.181 | -1.998(0.375) | -0.0005 | 0.140 | -1.979(0.205) | -0.010 | 0.042 |
| $\beta_{21}$ | -2.000 | -1.947(0.487) | -0.026 | 0.225 | -1.980(0.368) | -0.009 | 0.133 | -1.968(0.380) | -0.015 | 0.145 | -2.005(0.217) | 0.002 | 0.046 |
| $\beta_{22}$ | -2.000 | -1.881(0.982) | -0.059 | 0.919 | -1.851(0.505) | -0.074 | 0.273 | -2.031(0.436) | 0.015 | 0.190 | -1.987(0.257) | -0.006 | 0.066 |
| $d_{11}$ | 1.000 | 5.540(0.539) | 4.540 | 20.886 | 3.770(0.330) | 2.770 | 7.783 | 1.546(0.293) | 0.546 | 0.383 | 1.295(0.203) | 0.295 | 0.128 |
| $d_{12}$ | 0.500 | -0.017(0.280) | -1.034 | 0.341 | 0.075(0.184) | -0.849 | 0.213 | 0.342(0.173) | -0.314 | 0.054 | 0.427(0.100) | -0.144 | 0.015 |
| $d_{22}$ | 1.000 | 3.615(0.284) | 2.615 | 6.914 | 2.464(0.161) | 1.464 | 2.170 | 1.177(0.184) | 0.177 | 0.065 | 1.072(0.111) | 0.072 | 0.017 |
| $\sigma_e^2$ | 1.000 | 4.050(0.886) | 3.050 | 10.043 | 3.983(0.452) | 2.983 | 9.101 | 0.925(0.584) | -0.074 | 0.345 | 0.853(0.298) | -0.146 | 0.110 |
| $\delta_e$ | 3.000 | - | - | - | - | - | - | 2.904(0.489) | -0.031 | 0.248 | 3.022(0.180) | 0.007 | 0.032 |
| $\rho_1$ | 1.000 | 0.515(0.223) | -0.484 | 0.281 | 0.627(0.231) | -0.372 | 0.191 | 0.807(0.255) | -0.192 | 0.102 | 0.846(0.176) | -0.153 | 0.054 |
| $\rho_2$ | 2.000 | 1.908(0.126) | -0.045 | 0.023 | 2.004(0.177) | 0.001 | 0.030 | 2.118(0.203) | 0.059 | 0.055 | 2.095(0.118) | 0.047 | 0.022 |

*$d_{11}$, $d_{12}$ and $d_{22}$ are distinct elements of the matrix $\boldsymbol{D}$.*

**Table 3**: Result of simulation study for 500 samples under two distribution assumptions with 50% rate of random dropout. (Est.: posterior mean, S.E.: standard error, Rel. Bias: relative bias and MSE: mean square error).

| Model | | Normal model | | | | | | Skew-normal model | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Sample size | | $n = 100$ | | | $n = 500$ | | | $n = 100$ | | | $n = 500$ | | |
| parameters | real | Est.(S.E.) | Rel. Bias | MSE | Est.(S.E.) | Rel. Bias | MSE | Est.(S.E.) | Rel. Bias | MSE | Est.(S.E.) | Rel. Bias | MSE |
| $\beta_{11}$ | 10.000 | 10.004(0.320) | 0.0004 | 0.101 | 10.114(0.193) | 0.011 | 0.049 | 10.028(0.235) | 0.002 | 0.055 | 10.012(0.128) | 0.001 | 0.016 |
| $\beta_{12}$ | -3.000 | -3.149(0.299) | 0.049 | 0.110 | -3.091(0.140) | 0.030 | 0.027 | -3.031(0.156) | 0.010 | 0.025 | -3.012(0.089) | 0.004 | 0.007 |
| $\beta_{13}$ | -2.000 | -2.003(0.607) | 0.001 | 0.362 | -2.048(0.398) | 0.024 | 0.155 | -1.992(0.435) | -0.003 | 0.187 | -1.964(0.267) | -0.017 | 0.071 |
| $\beta_{21}$ | -3.000 | -3.137(0.658) | 0.045 | 0.444 | -2.924(0.301) | -0.025 | 0.093 | -3.002(0.345) | 0.0007 | 0.118 | -3.001(0.202) | 0.0002 | 0.040 |
| $\beta_{22}$ | -2.000 | -2.012(0.696) | 0.006 | 0.476 | -2.045(0.486) | 0.022 | 0.231 | -1.969(0.519) | -0.015 | 0.269 | -2.005(0.337) | 0.002 | 0.111 |
| $d_{11}$ | 1.000 | 5.761(0.553) | 4.760 | 22.968 | 3.881(0.356) | 2.881 | 8.427 | 1.607(0.346) | 0.607 | 0.488 | 1.298(0.229) | 0.298 | 0.141 |
| $d_{12}$ | 0.500 | -0.254(0.323) | -1.509 | 0.672 | -0.059(0.247) | -1.118 | 0.372 | 0.341(0.173) | -0.318 | 0.055 | 0.399(0.108) | -0.200 | 0.021 |
| $d_{22}$ | 1.000 | 3.924(0.363) | 2.924 | 8.684 | 2.529(0.181) | 1.529 | 2.372 | 1.216(0.181) | 0.216 | 0.079 | 1.093(0.106) | 0.093 | 0.019 |
| $\sigma_e^2$ | 1.000 | 4.042(0.710) | 3.042 | 9.753 | 3.894(0.421) | 2.894 | 8.552 | 0.928(0.638) | -0.072 | 0.409 | 0.834(0.259) | -0.166 | 0.093 |
| $\delta_e$ | 3.000 | - | - | - | - | - | - | 2.906(0.476) | -0.031 | 0.233 | 3.017(0.154) | 0.005 | 0.023 |
| $\rho_1$ | 1.000 | 0.737(0.287) | -0.263 | 0.150 | 0.672(0.192) | -0.327 | 0.142 | 0.826(0.326) | -0.173 | 0.135 | 0.882(0.151) | -0.117 | 0.036 |
| $\rho_2$ | 2.000 | 1.746(0.345) | -0.126 | 0.181 | 1.928(0.172) | -0.036 | 0.033 | 2.117(0.237) | 0.058 | 0.069 | 2.084(0.137) | 0.042 | 0.025 |

$d_{11}$, $d_{12}$ and $d_{22}$ are distinct elements of the matrix $\boldsymbol{D}$.

**Table 4**: Bayesian parameter estimates, posterior means (standard deviations, s.d.), and 95% HPDs for analysing the HIV data set using skew-normal distribution (Skew-normal model) and the normal distribution for error (Normal model).

| | Skew-normal model | | Normal model | |
|---|---|---|---|---|
| parameters | mean(s.d.) | 95% HPD | mean(s.d.) | 95% HPD |
| Intercept ($\beta_{11}$) | 10.351(0.591) | (9.186,11.589) | 10.589(0.713) | (9.212,11.921) |
| Time ($\beta_{12}$) | -0.362(0.048) | (-0.461,-0.274) | -0.352(0.052) | (-0.451,-0.241) |
| Time × Drug ($\beta_{13}$) | 0.019(0.071) | (-0.119,0.148) | 0.023(0.072) | (-0.119,0.172) |
| Gender ($\beta_{14}$) | -0.021(0.583) | (-1.209,1.198) | -0.255(0.673) | (-1.506,1.093) |
| PrevOI ($\beta_{15}$) | -4.516(0.429) | (-5.459,-3.749) | -4.642(0.517) | (-5.704,-3.661) |
| Stratum ($\beta_{16}$) | -0.194(0.422) | (-0.974,0.651) | -0.127(0.451) | (-0.938,0.818) |
| Intercept ($\beta_{21}$) | -4.769(0.699) | (-6.168,-3.402) | -4.767(0.713) | (-5.975,-3.171) |
| Drug ($\beta_{22}$) | 0.423(0.296) | (-0.141,1.003) | 0.373(0.292) | (-0.213,0.939) |
| Gender ($\beta_{23}$) | -0.498(0.407) | (-1.292,0.326) | -0.295(0.449) | (-1.146,0.568) |
| PrevOI ($\beta_{24}$) | 2.271(0.358) | (1.603,2.992) | 2.230(0.389) | (1.502,2.964) |
| Stratum ($\beta_{25}$) | 0.098(0.281) | (-0.441,0.674) | 0.076(0.288) | (-0.467,0.626) |
| $\rho_1$ | -0.314(0.044) | (-0.405,-0.229) | -0.299(0.042) | (-0.381,-0.216) |
| $\rho_2$ | -3.722(0.448) | (-4.711,-2.903) | -3.865(0.422) | (-4.633,-3.026) |
| $d_{11}$ | 15.281(1.195) | (13.078,17.710) | 16.131(1.192) | (13.940,18.591) |
| $d_{12}$ | -0.033(0.155) | (-0.340,0.271) | -0.040(0.156) | (-0.346,0.262) |
| $d_{22}$ | 0.468(0.037) | (0.399,0.547) | 0.472(0.039) | (0.401,0.554) |
| $\delta_e$ | 2.674(0.422) | (1.457,3.062) | – | – |
| $\sigma_e^2$ | 0.706(0.570) | (0.071,2.244) | 3.052(0.175) | (2.742,3.410) |
| $h_1$ | 0.146(0.088) | (0.038,0.376) | 0.119(0.077) | (0.028,0.325) |
| $h_2$ | 0.519(0.271) | (0.158,1.183) | 0.439(0.244) | (0.127,1.082) |
| $h_3$ | 1.397(0.706) | (0.431,3.142) | 1.207(0.643) | (0.369,2.881) |
| $h_4$ | 1.819(0.931) | (0.545,4.069) | 1.623(0.862) | (0.488,3.827) |
| | Model Comparison Criteria | | | |
| DIC | 6734.76 | | 9425.89 | |
| EAIC | 5290.922 | | 8810.331 | |
| EBIC | 5306.314 | | 8165.494 | |
| LPML | -3004.761 | | -3952.432 | |

*$d_{11}$, $d_{12}$ and $d_{22}$ are distinct elements of the matrix $\boldsymbol{D}$.*