Lifetime Data Analysis (2020) 26:339–368
https://doi.org/10.1007/s10985-019-09478-w

Quantile regression-based Bayesian joint modeling analysis of longitudinal–survival data, with application to an AIDS cohort study

Hanze Zhang[1] · **Yangxin Huang**[1]

Received: 21 December 2016 / Accepted: 23 May 2019 / Published online: 28 May 2019

Abstract
In longitudinal studies, it is of interest to investigate how repeatedly measured markers are associated with time to an event. Joint models have received increasing attention on analyzing such complex longitudinal–survival data with multiple data features, but most of them are mean regression-based models. This paper formulates a quantile regression (QR) based joint models in general forms that consider left-censoring due to the limit of detection, covariates with measurement errors and skewness. The joint models consist of three components: (i) QR-based nonlinear mixed-effects Tobit model using asymmetric Laplace distribution for response dynamic process; (ii) nonparametric linear mixed-effects model with skew-normal distribution for mismeasured covariate; and (iii) Cox proportional hazard model for event time. For the purpose of simultaneously estimating model parameters, we propose a Bayesian method to jointly model the three components which are linked through the random effects. We apply the proposed modeling procedure to analyze the Multicenter AIDS Cohort Study data, and assess the performance of the proposed models and method through simulation studies. The findings suggest that our QR-based joint models may provide comprehensive understanding of heterogeneous outcome trajectories at different quantiles, and more reliable and robust results if the data exhibits these features.

Keywords Asymmetric Laplace distribution · Bayesian inference · Nonlinear longitudinal quantile regression · Longitudinal–survival joint model · Limit of detection · Covariate measurement errors

✉ Yangxin Huang
yhuang@health.usf.edu

Hanze Zhang
hzhang1@health.usf.edu

[1] Department of Epidemiology and Biostatistics, College of Public Health, University of South Florida, Tampa, FL 33612, United States of America

1 Introduction

Data collected in many longitudinal medical studies record much information, not only repeated measures of disease biomarkers, but also survival status and follow-up time of patients. For example, in HIV/AIDS studies, viral load (the number of copies of HIV-1 RNA) and CD4 cell counts are important biomarkers of the severity of viral infection, disease progression, and treatment evaluation. In addition to the baseline measures, their longitudinal trends may be predictive of the risk of a terminal event (e.g., a steadily decreasing CD4 accounts may predict adverse events for HIV patients). Recently, joint models (JM) have been developed to assess the longitudinal process, survival process, and the relationship between them simultaneously, by sharing the random effects and/or related covariates (Brown 2009; Brown and Ibrahim 2003; Brown et al. 2005; Henderson et al. 2000; Rizopoulos 2011, 2012; Tsiatis and Davidian 2004; Wang and Taylor 2001). It is an active area of statistics and biostatistics research, because of its capability on the bias reduction and estimates' efficiency improvement, compared to separate modeling approaches (Chen et al. 2014; Huang et al. 2011; Wu et al. 2010).

The majority of longitudinal modeling methods, both linear and non-linear models, are typically based on mean regression, which concentrates on the average effect of response variable conditional on covariates (Huang and Dagne 2011; Liu and Wu 2007; Wu 2002; Wu et al. 2010; Yi et al. 2011). However, the mean trajectory of the longitudinal outcome may not always be of interest. In fact, in HIV/AIDS research, the effects of the covariates (e.g., CD4 counts) are more important on the left tail of the distribution where subjects are at higher risk. Additionally, if obvious outliers and/or heavy tails exist, mean regression may result in non-robust parameter estimates. For example, in the Multicenter AIDS Cohort Study (MACS) (Kaslow et al. 1987) (refer to Sect. 2.1 for the details of this study and data description), the viral load measurements are often highly skewed, even after $\log_{10}$ transformation (Fig. 1a). An effective solution is switching the concentration of mean regression to conditional quantiles (Davino et al. 2013; Koenker 2005; Koenker and Bassett Jr 1978). In contrast to the mean regression model focusing only on the central effects of the covariates, quantile regression (QR) model belongs to a robust model family, which can give a full scan of covariate effect at different quantiles of the outcome. Longitudinal QR models, specifically QR-based linear mixed-effects (LME) models have been proposed via different statistical approaches (He et al. 2003), including penalized least squares method (Koenker 2004), Barrodale–Roberts algorithm (Wang and Fygenson 2009), Expectation–Maximization (EM) algorithm (Farcomeni 2012), Monte Carlo Expectation–Maximization (MCEM) algorithm (Farcomeni and Viviani 2015; Ganjali and Baghfalaki 2015; Geraci and Bottai 2007; Liu and Bottai 2009), and Bayesian approach by Markov chain Monte Carlo (MCMC) procedure (Huang and Chen 2016; Huang 2016; Kim and Yang 2012; Luo et al. 2012; Tian and Tian 2015; Yuan and Yin 2010).

However, due to some inherent data features, modeling such complicated longitudinal data has many challenges. First, as a result of the low sensitivity of current standard assays, the measurement of viral load in an HIV positive individual is only accurate above a particular limit of detection (LOD), which is left-censored. For example, in some HIV/AIDS studies, values of HIV viral load below the LOD (40 copies/ml)

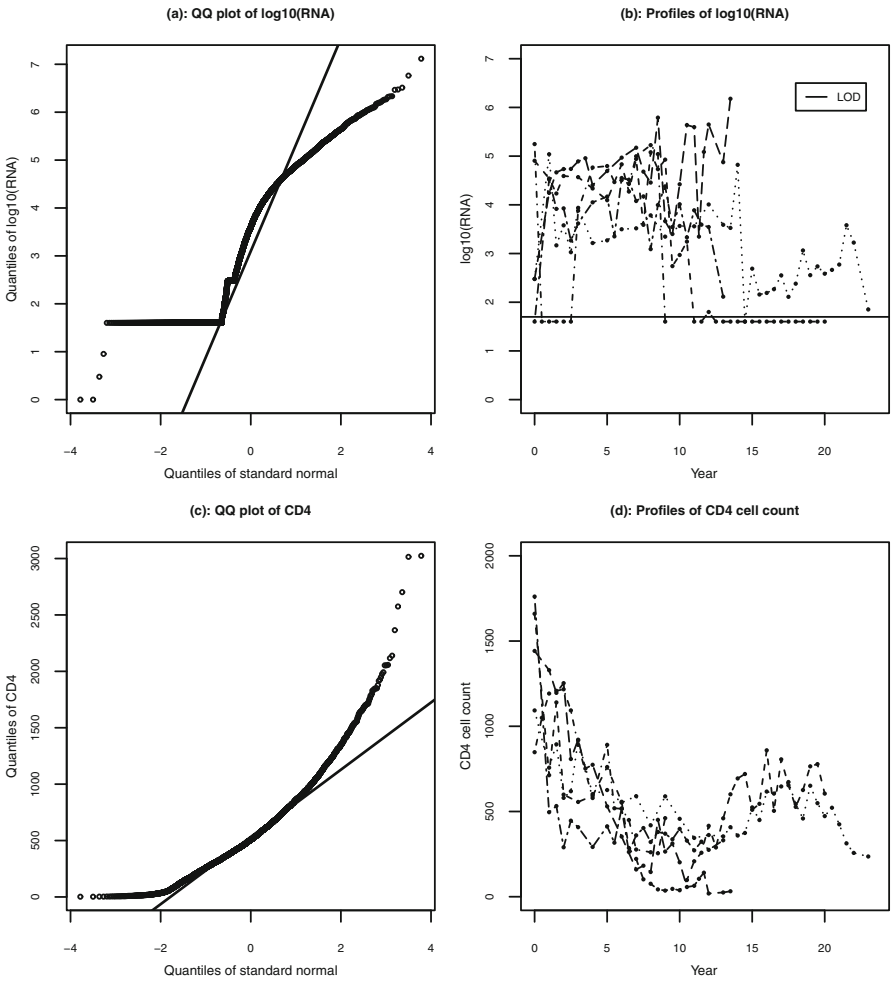

Fig. 1 QQ plots (left panel) of viral load (response) in $\log_{10}$ scale and CD4 covariate in plasma for 435 subjects with seroconversion from MACS data. Profiles (right panel) of viral load (response) in $\log_{10}$ scale and CD4 cell count (covariate) for for five representative patients

cannot be reliably quantified. Among subjects in MACS data, 26% observations of viral load fell below the LOD, which may lead to significant biases (Dagne and Huang 2011; Tian and Tian 2015). From Fig. 1b, it is obvious that some participants' viral load are below LOD. Dealing with censored observations, a simple but rough method is directly omitting or imputing the censored values by either the LOD or some arbitrary values, such as LOD/2. Instead, in order to get more accurate estimates, we adopt a Tobit model (Dagne and Huang 2011, 2012) to solve the left-censoring issue. Second, most of the studies assumed a normal distribution for the error terms of their models because of the mathematical tractability and computational convenience (Liu and Wu 2007; Wu 2002; Wu et al. 2010; Yi et al. 2011). In fact, Fig. 1c displays Q–Q plot

of the observed CD4 cell counts for subjects from the MACS data set. It is seen that CD4 cell counts are also highly skewed. The violation of the "normality" assumption could trigger biased or misleading inference (Huang and Dagne 2011; Huang et al. 2011). Compared to symmetric (normal) distribution, an asymmetric distribution such as skew-normal (SN) distribution (see Appendix A.1 in detail) should be more appropriate in modeling such "skewed" data (Arellano-Valle et al. 2007; Arellano-Valle and Genton 2005; Azzalini and Capitanio 1999; Sahu et al. 2003). Third, measurement error in covariate is another typical problem in longitudinal studies. For example, in AIDS studies, CD4 cell counts are often measured with substantial errors, and ignoring this phenomenon may lead to biased inference.

To the best of our knowledge, very few studies have done the QR-based JM for longitudinal–survival data with multiple data features, including LOD, skewed covariate with measurement errors. Statistical inference and analysis complicate dramatically when these issues exist simultaneously. In this article, thus, we propose a Bayesian approach for a JM consisting of (i) QR-based nonlinear mixed-effects Tobit (QR-NLMET) models with the asymmetric Laplace distribution (ALD) (Yu and Zhang 2005) (see Appendix A.2 in detail) for outcome process; (ii) nonparametric linear mixed-effects (LME) model with SN distribution for covariate process; and (iii) Cox proportional hazard model for survival process, which is linked to the outcome process through the random-effects. Approaching the complicated QR-NLMET JM from Bayesian perspective is more nature and straightforward, since it avoids many complicated approximations required by the frequentist methods.

The rest of the paper is organized as follows. In Sect. 2, we describe the data set that motivated this study and propose a specific JM. Section 3 presents the related Bayesian inferential method. In Sect. 4, we apply the proposed JM to the data set described in Sect. 2 and report the results. Section 5 conducts limited simulation studies to evaluate the performance of the proposed models and method. Finally, general discussion and conclusion is presented in Sect. 6.

2 A QR-based joint models and Bayesian inferential procedure

2.1 Motivating data set

The data set that motivated this research is from the MACS, a prospective study of the natural and treated histories of HIV-1 infection in homosexual and bisexual men conducted on multiple sites located in Baltimore, Chicago, Pittsburgh, and Los Angeles. The study participants have baseline and semiannual follow-up visits. The subset of MACS participants who seroconverted with HIV are of interest, since they were followed up from the time when they first developed antibodies to HIV. Many variables were collected during their follow-up which are used in our study, including patients demographic data (race and age at seroconversion), and laboratory results (viral load and CD4 counts). Researchers of the MACS conduct periodic death registry searches to obtain death information on all participants, even including those lost to follow up. Thus, survival information for each patient is also recorded, including survival status and the corresponding survival time. Due to the study design, survival time is

recorded at the year level, which is referred to as interval-censored death times (we discuss this in detail in Sect. 2.4). Finally, 435 patients with recorded HIV seroconversion, complete data in HIV viral load (at least three measurements), and CD4 counts, are included in our analysis. Of the 435 subjects with total 6437 observations, 1680 (26%) viral load fell below the LOD, and 203 (46.7%) subjects died during the study period.

Certain variables are transformed or imputed to benefit our analysis process. A $\log_{10}$ transformation of viral load, standardized CD4 cell counts (i.e. each CD4 count is subtracted by mean 556.25 and divided by standard deviation 325.01) and age at seroconversion are used in the analysis, in order to stabilize the variation of the measurements and to accelerate the convergence of estimation algorithm. As well as the survival time, one limitation of the public MACS data is that time of visit for each patient was only recorded at the year level. To account for the limitation, visit time is imputed according to the following criteria: (1) if a participant visited twice in year X, the time of the first visit is imputed as X, with the second visit as $X + 0.5$; (2) for a subject with three visits, time is imputed as X, $X + 0.33$, and $X + 0.67$, respectively. Additionally, in order to avoid unstable estimates with too small or large value, we rescale the original time t, so the new time ranges between 0 and 1.

2.2 Nonparametric skew-normal measurement error model for CD4 covariate

As discussed previously, the CD4 cell counts are usually measured with substantial errors and highly skewed. Various covariate models have been investigated in the literature, but most of them assume normal distribution for error term, which may lose the robustness against departures from normality in practice (Carroll et al. 2006; Liu and Wu 2007; Wu 2002). Moreover, the majority of the covariate models have been developed using parametric regression methods. Although such parametric models enjoy the simplicity, they suffer inflexibility when dealing with complicated longitudinal measures. Nonparametric regression models show the benefit to capture such complicated relationship by using flexible function forms. Therefore, in the presence of measurement error and non-normality in covariate, we employ a flexible empirical nonparametric mixed-effects model with SN distribution to quantify the covariate process. Denote the number of subjects by n and the number of measurements on the ith subject by n_i. Let z_{ij} be the observed CD4 covariate value for individual i at time t_{ij} $(i = 1, \ldots n;\ j = 1, \ldots, n_i)$.

$$\begin{aligned} z_{ij} &= w(t_{ij}) + h_i(t_{ij}) + \epsilon_{ij} \ \ (\equiv z^*_{ij} + \epsilon_{ij}), \\ \boldsymbol{\epsilon}_i &\overset{\text{iid}}{\sim} SN_{n_i}\left(-\sqrt{2/\pi}\delta \mathbf{1}_{n_i}, \sigma_1^2 \boldsymbol{I}_{n_i}, \delta \boldsymbol{I}_{n_i}\right), \end{aligned} \tag{1}$$

where $\boldsymbol{z}_i = (z_{i1}, \ldots, z_{in_i})^T$ with z_{ij} being the observed CD4 covariate value for individual i at time t_{ij}; $\boldsymbol{z}^*_i = (z^*_{i1}, \ldots, z^*_{in_i})^T$ and $z^*_{ij} = w(t_{ij}) + h_i(t_{ij})$ may be viewed as the true (but unobservable) covariate value at time t_{ij}; $w(t_{ij})$ and $h_i(t_{ij})$ are unknown nonparametric smooth fixed-effects and random-effects functions, respectively; and $\boldsymbol{\epsilon}_i = (\epsilon_{i1}, \ldots, \epsilon_{in_i})^T$ follows a multivariate SN distribution with unknown

variance parameter σ_1^2 and skewness parameter δ, where the location parameter is set as $-\sqrt{2/\pi}\delta \mathbf{1}_{n_i}$ in order to have a zero mean vector for error term and $\mathbf{1}_{n_i} = (1, \ldots, 1)^T$. It is noticed that we specify the skewness matrix being $\delta \boldsymbol{I}_{n_i}$, indicating that we are interested in skewness of overall pooled CD4 measurements from all the patients.

Compared to the parametric LME model, the nonparametric LME model (1) is more flexible. The fixed smooth function $w(t)$ represents population average of the covariate process, while the random smooth function $h_i(t)$ is used to incorporate the inter-individual variation in the covariate process. We assume that $h_i(t)$ is the realization of a zero-mean stochastic process. To fit model (1), a regression spline method is applied to $w(t)$ and $h_i(t)$, details can be found in literature (Davidian and Giltinan 1995; Wu and Zhang 2006). Briefly, the working principle of regression spline is described as: a linear combination of spline basis functions is used to approximate $w(t)$ and $h_i(t)$. For example, $w(t)$ and $h_i(t)$ can be approximated by a linear combination of basis functions $\boldsymbol{\Psi}_p(t) = \{\psi_0(t), \psi_1(t), \ldots, \psi_{p-1}(t)\}^T$ and $\boldsymbol{\Phi}_p(t) = \{\phi_0(t), \phi_1(t), \ldots, \phi_{p-1}(t)\}^T$, respectively. That is,

$$\begin{aligned} w(t) &\approx w_p(t) = \sum_{l=0}^{p-1} \alpha_l \psi_l(t) = \boldsymbol{\Psi}_p(t)^T \boldsymbol{\alpha}, \\ h_i(t) &\approx h_{iq}(t) = \sum_{l=0}^{q-1} a_{il} \phi_l(t) = \boldsymbol{\Phi}_q(t)^T \boldsymbol{a}_i, \end{aligned} \tag{2}$$

where $\boldsymbol{\alpha} = (\alpha_0, \ldots, \alpha_{p-1})^T$ is a $p \times 1$ vector of fixed-effects, and $\boldsymbol{a}_i = (a_{i0}, \ldots, a_{i,q-1})^T$ $(q \leq p)$ is a $q \times 1$ vector of random-effects. We assume that $\boldsymbol{a}_i \overset{\text{iid}}{\sim} N_l(\mathbf{0}, \boldsymbol{\Sigma}_a)$ with $\boldsymbol{\Sigma}_a$ being unrestricted covariance matrix. Based on the assumption of $h_i(t)$, $\boldsymbol{a}_i$ is regarded as iid realizations of a zero-mean random vector. We consider natural cubic spline bases with the percentile-based knots in this model. Selection of optimal degree of regression spline and numbers of knots, in other words, the optimal sizes of p and q, is determined according to the Akaike information criterion (AIC) or the Bayesian information criterion (BIC) (Davidian and Giltinan 1995).

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

$$\begin{aligned} z_{ij} &\approx \boldsymbol{\Psi}_p(t_{ij})^T \boldsymbol{\alpha} + \boldsymbol{\Phi}_q(t_{ij})^T \boldsymbol{a}_i + \epsilon_{ij} \approx z_{ij}^* + \epsilon_{ij}, \\ \boldsymbol{\epsilon}_i &\overset{\text{iid}}{\sim} SN_{n_i}\left(-\sqrt{2/\pi}\delta \mathbf{1}_{n_i}, \sigma_1^2 \boldsymbol{I}_{n_i}, \delta \boldsymbol{I}_{n_i}\right) \end{aligned} \tag{3}$$

2.3 The QR-NLMET models for viral response

We use y_{ij} to denote the $\log_{10}$-transformation of the response variable, viral load, for the ith subject at time t_{ij} $(i = 1, 2, ..., n,\ j = 1, 2, ..., n_i)$. To introduce the Tobit model accounting for the left-censoring data in our JM framework, we denote the

observed value y_{ij} by (q_{ij}, c_{ij}), where c_{ij} is the censoring indicator and q_{ij} is the latent response variable. The latent q_{ij} is observed, as y_{ij}, then $c_{ij} = 0$, if and only if $y_{ij} > \zeta$ (a known constant LOD, $\log_{10}(40) = 1.602$ in our data example). Otherwise, y_{ij} is treated as missing value, and $c_{ij} = 1$.

Viral dynamic models can be formulated through a system of ordinary differential equations (ODE) (Huang et al. 2006; Wu and Ding 1999). Wu and Ding (1999) provided some biological arguments for a two-compartment viral dynamic model where two phases of viral decay are discussed. Under some reasonable assumptions, a simplified approximation of ODE has been proposed as follows to capture the viral load trajectory.

$$y(t) = \log_{10}(e^{p_1 - \lambda_1 t} + e^{p_2 - \lambda_2 t}), \tag{4}$$

where $y(t)$ is the $\log_{10}$ scaled viral load levels at time t. λ_1 and λ_2 are the first- and second-phase viral decay rates, which may represent the minimum turnover rate of productively infected cells and that of latently or long-lived infected cells, respectively (Perelson et al. 1997). The parameters p_1 and p_2 are macro-parameters, and $e^{p_1} + e^{p_2}$ represents the baseline viral load at time $t = 0$. It is generally assumed that $\lambda_1 > \lambda_2$, which assures that the model is identifiable and is appropriate for empirical studies (Wu and Ding 1999). From the Fig. 1b, we noted that the trajectories of viral load are dramatically different. For some patients, viral loads decrease at the beginning and then increase indicating a viral rebound; but the viral load responses of some other patients decrease over time and stay at a relatively low level. It suggests that the second-phase viral decay rate λ_2 may vary over time, and variation in the dynamic parameters, particularly λ_2, may be partially associated with time-varying covariates such as repeated CD4 cell counts. Thus, it may not be reasonable to assume that second-phase viral decay rate λ_2 is a constant since the viral load varies at the later stage during a long-term treatment. To model the complicated long-term HIV dynamics, a natural extension of Eq. (4) assuming that the second-phase viral decay rate λ_2 changes over time, is applied, which is a function of time-varying CD4 counts. Thus, we determine to apply the following QR-NLMET model for the viral load response.

$$\begin{aligned} y_{ij} &= \log_{10}(e^{p_{i1} - \lambda_{i1} t_{ij}} + e^{p_{i2} - \lambda_{ij2} t_{ij}}) + e_{ij} \\ p_{i1} &= \beta_1 + b_{i1}, \quad \lambda_{i1} = \beta_2 + \beta_3 z_{i0} + b_{i2}, \\ p_{i2} &= \beta_4 + b_{i3}, \quad \lambda_{ij2} = \beta_5 + \beta_6 z^*_{ij} + b_{i4}, \end{aligned} \tag{5}$$

where z_{i0} and z^*_{ij} are the baseline CD4 counts and true (but unobservable) value of CD4 covariate for the ith subject at time t_{ij}, respectively; $\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_6)^T$ and $\boldsymbol{\beta}_{ij} = (p_{i1}, p_{i2}, \lambda_{i1}, \lambda_{ij2})^T$ are population parameters and individual-specific parameters for the ith subject at time t_{ij}, respectively; $\boldsymbol{e}_i = (e_{i1}, \ldots, e_{in_i})^T$ represents vector of within-individual random error e_{ij} which follows $ALD(0, \sigma, \tau)$ whose distribution is restricted to have the τth quantile equal to zero with the scale parameter $\sigma\,(> 0)$, i.e., $Q_{e_{ij}}(\tau | \boldsymbol{b}_i; t_{ij}, z_{ij}, z_{i0}) = 0$, where $Q_{e_{ij}}(\tau|\cdot) \equiv F^{-1}_{e_{ij}}(\tau|\cdot)$ is the inverse of cumulative distribution function (cdf) of e_{ij} evaluated at τ with $0 < \tau < 1$; $\boldsymbol{b}_i = (b_{i1}, \ldots, b_{i4})^T$ is random-effects following $N_4(\boldsymbol{0}, \boldsymbol{\Sigma}_b)$ with $\boldsymbol{\Sigma}_b$ being unrestricted covariance matrix.

2.4 Survival model for time-to-death data

Cox proportional hazard model is adopted for time-to-death process. We assume that the distribution of T_i, the death time for the ith subject, depends on the random-effects $\boldsymbol{b}_i$, representing individual-specific longitudinal processes, and other survival covariates $\boldsymbol{x}_i$, respectively. We therefore consider a frailty model for T_i, which is linked to QR-NLMET model (5) through the random-effects $\boldsymbol{b}_i$. Additional covariates $\boldsymbol{x}_i$ are assumed to be associated with the risk of death but not with the longitudinal measurements. Specifically, in the survival process, the conditional hazard rate of T_i at time t_i is expressed as

$$\lambda(t_i|\boldsymbol{b}_i, \boldsymbol{x}_i) = \lambda_0(t_i)\exp(\boldsymbol{\gamma}_1^T\boldsymbol{b}_i + \boldsymbol{\gamma}_2^T\boldsymbol{x}_i) = \lambda_0(t_i)\exp(\boldsymbol{\gamma}^T\boldsymbol{d}_i), \tag{6}$$

where $\lambda_0(t_i)$ is the baseline hazard function, $\boldsymbol{d}_i = (\boldsymbol{b}_i^T, \boldsymbol{x}_i^T)^T$, $\boldsymbol{\gamma} = (\boldsymbol{\gamma}_1^T, \boldsymbol{\gamma}_2^T)^T$, $\boldsymbol{\gamma}_1$ and $\boldsymbol{\gamma}_2$ are unknown parameters linking the random-effects $\boldsymbol{b}_i$ and covariates $\boldsymbol{x}_i$ to the conditional hazard rate, respectively.

Let $\boldsymbol{\rho}_i = (\rho_{i1}, \ldots, \rho_{in_i})^T$ be a vector of censoring ('death') indicator for individual i: $\rho_{ij} = 1$ if the patient has died by time t_{ij}, otherwise, $\rho_{ij} = 0$. We assume that $\rho_{i1} = 0$ for all subjects. In the MACS example, for individual i, let T_i be the time to death and assume $P(T_i < \infty) = 1$, but it is not exactly recorded due to the limitation of study design. This type of survival time is usually called interval-censored death time. Specifically, if the death of individual i is recorded at time t_{ij}, it means that the true death time T_i takes place during the time period $(t_{i,(j-1)}, t_{ij}]$, that is, $t_{i,(j-1)} < T_i \le t_{ij}$. Also, we take $\rho_{i1} = \ldots = \rho_{i,(j-1)} = 0$ and $\rho_{ij} = \ldots = \rho_{in_i} = 1$. If no such event (death) has happened, we treat $\rho_{ij} = 0$ for $j = 1, \ldots, n_i$, and thus $T_i > t_{in_i}$.

From Eq. (6), the probability of time-to-death is given as

$$\begin{aligned} p_{ij} &= P(\rho_{ij} = 1|\rho_{ij^*} = 0, 0 \le j^* < j) = 1 - P(T_i \ge t_{ij}|T_i > t_{i,j-1}) \\ &= 1 - \frac{P(T_i \ge t_{ij})}{P(T_i \ge t_{i,j-1})} = 1 - exp\left(-\int_{t_{i,j-1}}^{t_{ij}} \lambda_0(t)dt\ exp(\boldsymbol{\gamma}^T\boldsymbol{d}_i)\right) \\ &= 1 - exp(-exp(\gamma_{0j} + \boldsymbol{\gamma}^T\boldsymbol{d}_i)), \end{aligned} \tag{7}$$

where $\gamma_{0j} = \log\left(\int_{t_{i,j-1}}^{t_{ij}} \lambda_0(t)dt\right)$, $j = 1, \ldots, n_i$. Given the current observation mechanism, only the finite number of parameters γ_{0j} are required to handle, instead of the unknown baseline hazard function $\lambda_0(t)$ in the parameter estimation-based method. The contribution to the likelihood from the time-to-death part for the ith subject is denoted by $f(\boldsymbol{\rho}_i|\boldsymbol{d}_i)$. Thus, we have

$$f(\boldsymbol{\rho}_i|\boldsymbol{d}_i) = \prod_{j=1}^{n_i} f(\rho_{ij}|\rho_{ij^*}, 0 \le j^* < j; \boldsymbol{d}_i), \tag{8}$$

where $f(\rho_{ij}|\rho_{ij^*}, 0 \le j^* < j; \boldsymbol{d}_i) = p_{ij}^{\rho_{ij}}(1 - p_{ij})^{1-\rho_{ij}}$ and ρ_{ij} equals to 0 before death and 1 after death. Let $v_i = max(t_{ij} : \rho_{ij} = 0)$ and $u_{ij} = min(t_{ij} : \rho_{ij} = 1)$.

With (8), the following probability of time-to-death can be expressed as

$$P(v_i < T_i \leq u_i|\boldsymbol{d}_i) = exp\left(-\int_0^{v_i} \lambda_0(t)dt\ exp(\boldsymbol{\gamma}^T\boldsymbol{d}_i)\right) \cdot \left[1 - exp\left(-\int_{v_i}^{u_i} \lambda_0(t)dt\ exp(\boldsymbol{\gamma}^T\boldsymbol{d}_i)\right)\right], \tag{9}$$

where $u_i = \infty$ if $\rho_{ij} = 0$ for $j = 1, \ldots, n_i$. The process we discussed above is an alternative way to approximate the Cox proportional hazard model through counting process (Clayton 1991).

3 Simultaneous Bayesian inferential approach

Estimation of a longitudinal–survival JM could be conducted in two general ways. The first estimation approach is based on the likelihood inferential methods, such as EM and MCEM algorithm (Farcomeni and Viviani 2015; Rizopoulos 2010). Although these methods for JM with multiple data feature may be favourable, the computational cost for such a complicated model is extremely intensive, and particularly, may lead to convergence problem (Tang et al. 2017; Wu et al. 2010). The second method is Bayesian inferential approach, which may be more natural to deal with such complex models, simpler to implement when prior information is available, and overcome the convergence problem. Thus, we propose a fully Bayesian framework for the JM, which consists of covariate measurement error model (1), QR-NLMET model (5), and Cox proportional hazard model (6), to simultaneously estimate parameters. We obtain numerical approximations to posterior distributions using MCMC procedure.

We assume that $\boldsymbol{\epsilon}_i$, $\boldsymbol{e}_i$, $\boldsymbol{a}_i$ and $\boldsymbol{b}_i$ are mutually independent of each other. In order to specify the JM for MCMC computation, by introducing the random vector $\boldsymbol{w}_i = (w_{i1}, \ldots, w_{in_i})^T$ and random variable v_{ij} based on the stochastic representations in Eqs. (A.6) and (A.11) presented in Appendix, we hierarchically formulate the joint models (1), (5), and (6) as follows

$$\begin{aligned}
&y_{ij}|z_{ij}, \boldsymbol{b}_i, \boldsymbol{a}_i, v_{ij} \sim N\left(g(t_{ij}, \boldsymbol{\beta}_{ij}) + \vartheta_1 v_{ij}, \vartheta_2\sigma v_{ij}\right),\\
&z_{ij}|\boldsymbol{a}_i, w_{ij} \sim N(z_{ij}^* + \delta(w_{ij} - \sqrt{2/\pi}), \sigma_1^2),\\
&\boldsymbol{a}_i \sim N_3(\boldsymbol{0}, \boldsymbol{\Sigma}_a),\ \boldsymbol{b}_i \sim N_4(\boldsymbol{0}, \boldsymbol{\Sigma}_b),\\
&w_{ij} \sim N(0,1)I(w_{ij} > 0),\ v_{ij} \sim Exp\left(\frac{1}{\sigma}\right),\\
&T_i \sim F(t_i|\boldsymbol{d}_i) = \int f(\boldsymbol{\rho}_i|\boldsymbol{d}_i),
\end{aligned} \tag{10}$$

where $g(t_{ij}, \boldsymbol{\beta}_{ij}) = \log_{10}(e^{p_{i1}-\lambda_{i1}t_{ij}} + e^{p_{i2}-\lambda_{ij2}t_{ij}})$ described in Eq. (5), $\vartheta_1 = (1 - 2\tau)/[\tau(1-\tau)]$ and $\vartheta_2 = 2/[\tau(1-\tau)]$; $I(w_{ij} > 0)$ is an indicator function and $w_{ij} \sim N(0,1)$ truncated in the space $w_{ij} > 0$ (i.e., standard half-normal distribution).

Denote $\boldsymbol{\theta} = \{\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \sigma_1^2, \sigma, \boldsymbol{\Sigma}_a, \boldsymbol{\Sigma}_b, \delta\}$ as the collection of unknown population parameters in the models (1), (5), and (6). Under the Bayesian framework, we specify prior distributions for all of these unknown parameters as follows

$$\begin{aligned} &\boldsymbol{\alpha} \sim N_3(\boldsymbol{\alpha}_0, \boldsymbol{\Omega}_1),\ \boldsymbol{\beta} \sim N_6(\boldsymbol{\beta}_0, \boldsymbol{\Omega}_2),\ \boldsymbol{\gamma} \sim N_4(\boldsymbol{\gamma}_0, \boldsymbol{\Omega}_3), \\ &\sigma \sim IG(\omega_1, \omega_2),\ \sigma_1^2 \sim IG(\omega_3, \omega_4), \\ &\boldsymbol{\Sigma}_a \sim IW(\boldsymbol{\Omega}_4, \omega_5),\ \boldsymbol{\Sigma}_b \sim IW(\boldsymbol{\Omega}_5, \omega_6),\ \delta \sim N(0, \omega_7), \end{aligned} \tag{11}$$

where the mutually independent Normal (N), Inverse Gamma (IG) and Inverse Wishart (IW) prior distributions are chosen to facilitate computations. The hyper-parameter matrices $\boldsymbol{\Omega}_k$ ($k = 1, \ldots, 5$) can be assumed to be diagonal for convenient implementation.

Let $f(\cdot)$, $f(\cdot|\cdot)$, $F(\cdot|\cdot)$ and $\pi(\cdot)$ denote a density function, a conditional density function, a cumulative density function (c.d.f) and a prior density function, respectively. We assume that $\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \sigma_1^2, \sigma, \boldsymbol{\Sigma}_a, \boldsymbol{\Sigma}_b$ and δ are independent of each other, then we have $\pi(\boldsymbol{\theta}) = \pi(\boldsymbol{\alpha})\pi(\boldsymbol{\beta})\pi(\boldsymbol{\gamma})\pi(\sigma_1^2)\pi(\sigma)\pi(\boldsymbol{\Sigma}_a)\pi(\boldsymbol{\Sigma}_b)\pi(\delta)$. After specifying the JM for the observed data and the prior distributions for the unknown parameters, we can make Bayesian statistical inference for the parameters based on their posterior distributions. Thus, the joint posterior density of $\boldsymbol{\theta}$ based on the observed data $\mathcal{D} = \{(y_{ij}, c_{ij}, z_{ij}, z_{i0}, x_i, \rho_{ij}), i = 1, \ldots, n;\ j = 1, \ldots, n_i\}$ can be given by

$$\begin{aligned} f(\boldsymbol{\theta}|\mathcal{D}) \propto \Bigg\{ &\prod_{i=1}^{n} \int \int \prod_{j=1}^{n_i} f(y_{ij}|z_{ij}, \boldsymbol{b}_i, \boldsymbol{a}_i, v_{ij})^{1-c_{ij}} F(\zeta|z_{ij}, \boldsymbol{b}_i, \boldsymbol{a}_i, v_{ij})^{c_{ij}} \\ &f(z_{ij}|\boldsymbol{a}_i, w_{ij}) f(\boldsymbol{\rho}_i|\boldsymbol{d}_i) f(\boldsymbol{b}_i) f(\boldsymbol{a}_i) f(v_{ij}) f(w_{ij}|w_{ij} > 0) d\boldsymbol{b}_i d\boldsymbol{a}_i \Bigg\} \pi(\boldsymbol{\theta}). \end{aligned} \tag{12}$$

Generally, the integrals in (12) are of high dimension and do not have a closed form. Analytic approximations to the integrals may not be sufficiently accurate. Therefore, it is prohibitive to directly calculate the posterior distribution of $\boldsymbol{\theta}$ based on the observed data. As an alternative, the MCMC procedure can be used to sample population parameters $\boldsymbol{\theta}$, and random-effects $\boldsymbol{a}_i$ and $\boldsymbol{b}_i$ ($i = 1, \ldots, n$), from conditional posterior distributions based on (12), by employing the Gibbs sampler along with the Metropolis-Hastings (M-H) algorithm. This process is repeated in iterations of MCMC procedure until convergence is reached. An important advantage of the above representations based on the JM is that they are easily implemented using the public and freely available WinBUGS software (Lunn et al. 2000) interacted with a function called bugs in a package named R2WinBUGS of R. Another advantage is that when WinBUGS software is used to implement our modeling approach, it is not necessary to explicitly specify the full conditional posterior distributions for parameters to be estimated. Although their derivations are straightforward by working the complete joint posterior (12), some cumbersome algebra will be involved. We, thus, omit them here to save space.

4 Application to MACS data

4.1 Model implementation

Section 2.1 has briefly described the MACS data set that motivated this research. As showed in Fig. 1d previously, the covariate CD4 cell counts described in Sect. 2 exhibited the asymmetric feature. It is therefore plausible to fit the data by considering covariate model error with an SN distribution. Furthermore, there are substantial observations below LOD in viral load. From the biological perspective, the longitudinal process and time-to-event process are inherently connected. Toward the end, the following two JM with specifying different distributions were employed to compare their performance:

- **Model SN:** e_{ij} and ϵ_{ij} follow ALD and SN distribution, respectively.
- **Model N:** e_{ij} and ϵ_{ij} follow ALD and normal distribution, respectively.

Since a normal distribution is a special case of an SN distribution when skewness parameter is zero, we investigate how the covariate model (1) using an asymmetric SN distribution in JM contributes to modeling results and parameter estimation in comparison with that using a symmetric normal distribution.

To empirically model the complicated CD4 process with non-negligible measurement errors, we apply a nonparametric mixed-effects model. Compared to the parametric models, this nonparametric method is flexible and works well for complex longitudinal data. We use linear combinations of natural cubic splines with percentile-based knots to approximate $w(t)$ and $h_i(t)$ discussed in (1). Following the study of Liu and Wu (2007), we set $\psi_0(t) = \phi_0(t) = 1$ and take the same natural cubic splines in the approximations (2) with the degree of regression spline $q \leq$ the numbers of knots p (in order to limit the dimension of random-effects). The values of p and q are determined according to the AIC/BIC criteria. The AIC/BIC values were evaluated in various models with $(p, q) = \{(1, 1), (2, 1), (2, 2), (3, 1), (3, 2), (3, 3)\}$. Among them, we found that the model with $(p, q) = (3, 3)$ has the smallest AIC/BIC value. Thus, we adopted the following nonparametric mixed-effects CD4 covariate model,

$$\begin{aligned} z_{ij} &= (\alpha_0 + a_{i0})\psi_0(t_{ij}) + (\alpha_1 + a_{i1})\psi_1(t_{ij}) + (\alpha_2 + a_{i2})\psi_2(t_{ij}) \\ &\quad +\epsilon_{ij} \quad (\equiv z^*_{ij} + \epsilon_{ij}) \end{aligned} \tag{13}$$

where z_{ij} is the observed CD4 value after standardization at time t_{ij}, $\psi_1(\cdot)$ and $\psi_2(\cdot)$ are two basis functions given in Sect. 2.2, $\boldsymbol{\alpha} = (\alpha_0, \alpha_1, \alpha_2)^T$ is a vector of population (fixed-effects) parameters, $\boldsymbol{a}_i = (a_{i0}, a_{i1}, a_{i2})^T$ is a vector of individual (random-effects) parameters, and ϵ_{ij} is the measurement error at time t_{ij} following an SN distribution.

Cox proportional hazard model is applied to model time-to-death process. As discussed in Sect. 2.4, we assume that the survival time T_i for the ith subject depends on the random-effects $\boldsymbol{b}_i$ and other survival covariates $\boldsymbol{x}_i$. Other baseline covariates include age at seroconversion (age_i), and race ($race_i$), with $race_i = 1$ if subject i is white, and $race_i = 0$ otherwise. To avoid extremely heavy computational load, we do not put random-effects $\boldsymbol{a}_i$ and other covariates into the model. Therefore, the Cox

submodel (6) is specified as

$$\lambda(t_i|\boldsymbol{b}_i, \boldsymbol{x}_i) = \lambda_0(t_i)\exp(\gamma_1 b_{i2} + \gamma_2 b_{i4} + \gamma_3 age_i + \gamma_4 race_i), \tag{14}$$

where $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_4)^T$ is the parameters corresponding to the random-effects (b_{i2}, b_{i4}), covariates $(age_i, race_i)$. It is noted that the random-effects b_{i2} and b_{i4} represent individual variations in the first- and second-phase viral load decay rates, respectively, which may be predictive for survival time. Since b_{i1} and b_{i3} represent variations in the baseline viral loads, which do not appear to be highly correlated to the risk of death, to reduce the number of parameters, they are excluded from model (14).

To specify the values for the hyper-parameters in the prior distributions (11), we take weakly informative prior distributions for the parameters. In particular, (i) fixed-effects are taken to be independent normal distribution $N(0, 100)$ for each element of the population parameter vectors $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, and $\boldsymbol{\gamma}$; (ii) we assume a noninformative inverse Gamma prior distribution $IG(0.01, 0.01)$, which has mean 1 and variance 100, for scale parameter σ and σ_1^2; (iii) the priors for the variance-covariance matrices of the random-effects $\boldsymbol{\Sigma_a}$ and $\boldsymbol{\Sigma_b}$ are taken to be inverse Wishart distributions $IW(\text{diag}(0.01, 0.01, 0.01), 3)$ and $IW(\text{diag}(0.01, 0.01, 0.01, 0.01), 4)$; (iv) for the skewness parameter δ, a normal distributions $N(0, 100)$ is chosen. Sensitivity analysis is also conducted.

The MCMC sampler was implemented using WinBUGS software (Lunn et al. 2000) interacted with R2WinBUGS of R software, and the program code is available from authors upon request. When the MCMC procedure was applied to the read data (MACS), convergence of the generated samples was assessed using standard tools within WinBUGS software such as trace plots and Gelman–Rubin (GR) diagnostics (Gelman and Rubin 1992). Figure 2 shows the trace plots, autocorrelation and dynamic version of GR diagnostic plots based on Model SN for the representative parameters $\alpha_2, \beta_3, \gamma_4$, and δ. We observe from trace plots (left panel) that the lines of three different chains mix or cross in trace, implying that convergence is reached. For the plots of GR diagnostics (middle panel) where the three curves are given: the middle and bottom curves below the dashed horizontal line (indicated by the value one) represent the pooled posterior variance ($\hat{V}$) and average within-sample variance ($\hat{W}$), respectively, and the top curve represents their ratio ($\hat{R}$). It is seen that $\hat{R}$ is generally expected to be higher than one at the initial stage of the algorithms, but $\hat{R}$ tends to 1, and $\hat{V}$ and $\hat{W}$ stabilize as the number of iterations increase, indicating that the algorithm has approached convergence. We further monitor convergence using autocorrelation plots (right panel) that autocorrelations are very low with a lag being 50, implying that convergence is ensured. When these criteria suggested the convergence of chains, we proposed that, after an initial number of 100,000 burn-in iterations of three chains of length 150,000, every 50th MCMC sample was retained from the next 50,000 for each chain. Thus, we obtained a total of 3000 samples of targeted posterior distributions of the unknown parameters for statistical inference. The Bayesian modeling approach in conjunction with the JM with different scenarios is applied to fit the MACS data described in Sect. 2 for inference. In the following section, we report the results based on the two scenarios highlighted above.

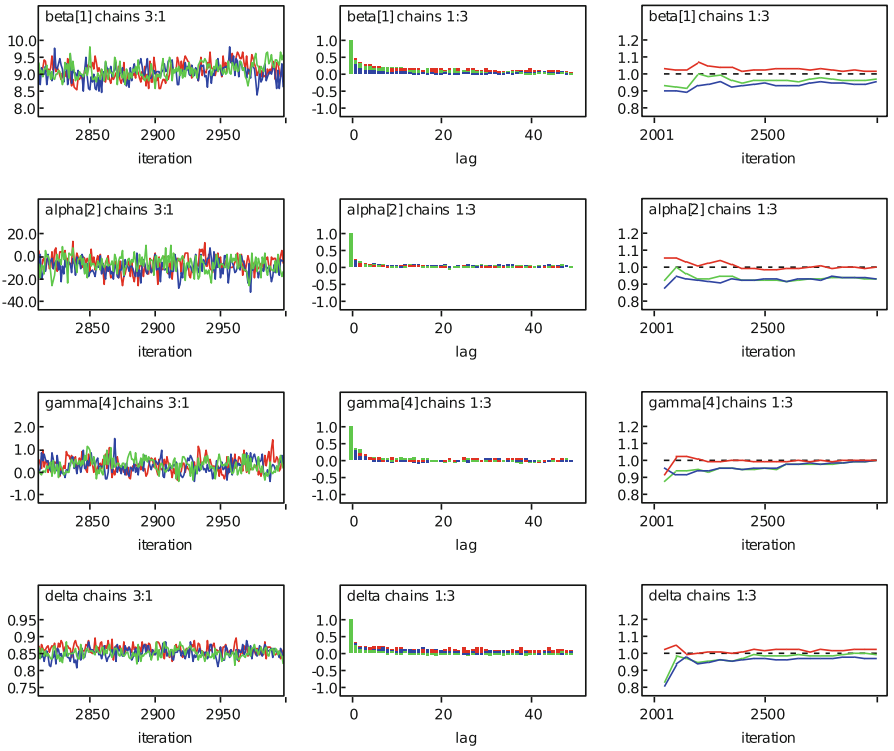

Fig. 2 Convergence diagnostics with three Markov chains as obtained from the WinBUGS software for representative parameters based on Model SN with $\tau = 0.5$ quantile: trace plots (left panel); autocorrelation plots (middle panel); Gelman-Rubin (GR) diagnostic plots (right panel), where the middle and bottom curves below the dashed horizontal line (indicated by the value one) represent the pooled posterior variance ($\hat{V}$) and average within-sample variance ($\hat{W}$), respectively, and the top curve above the dashed horizontal line represents their ratio ($\hat{R}$)

4.2 Data analysis results

Bayesian joint modeling approach in conjunction with the QR-NLMET JM at the five quantiles of $\tau = 0.05, 0.25, 0.50, 0.75$ and 0.95 is used to fit the viral load, CD4, and survival data jointly. Table 1 presents the population posterior mean (PM), the corresponding 95% credible interval (CI) for fixed-effects parameters based on the two proposed models (N and SN). The following findings are obtained for the results of estimated parameters.

According to the results from Models N and SN at five quantiles of $\tau = 0.05, 0.25, 0.50, 0.75$ and 0.95 in Table 1, interesting findings are discovered. First, as τ increases, due to the increase of β_1 and β_4 (except β_1 at $\tau = 0.95$), the baseline viral load, $e^{p_1} + e^{p_2}$, increases. Second, the estimates of β_2 and β_3, which are parameters associated with the first-phase viral decay rate, have the smallest value at median ($\tau = 0.50$), and increase as quantile goes extremely (in both directions) with the largest value at $\tau = 0.95$. It indicates that the first-phase viral decay rate increases

Table 1 Summary of estimated posterior mean (PM) of population (fixed-effects) parameters, the corresponding lower limit (L_{CI}) and upper limit (U_{CI}) of 95% equal-tail credible interval (CI) as well as DIC values

Method	Model		β_1	β_2	β_3	β_4	β_5	β_6	α_1	α_2	α_3	γ_1	γ_2	γ_3	γ_4	δ	DIC	EPD
$\tau = 0.05$																		
JM	N	PM	−43.92	20.13	−3.81	−45.97	5.69	3.00	−0.06	0.65	−64.92	−0.09	−0.03	0.14	0.18	–	81980.9	127.0
		L_{CI}	−47.89	12.89	−8.29	−49.74	0.22	0.82	−0.14	−14.36	−82.99	−1.36	−1.28	−0.03	−0.23	–		
		U_{CI}	−42.30	30.60	2.82	−43.75	9.71	5.02	0.03	15.60	−48.23	1.34	1.27	0.30	0.63	–		
JM	SN	PM	−53.91	26.01	−4.91	−56.58	8.33	4.75	−0.05	−7.26	−67.99	−0.12	−0.09	0.14	0.18	0.86	75486.9	70.7
		L_{CI}	−57.41	17.41	−8.89	−60.33	1.97	1.48	−0.12	−21.93	−85.91	−1.34	−1.40	−0.02	−0.23	0.83		
		U_{CI}	−52.02	37.05	2.07	−54.04	13.33	7.83	0.04	7.34	−50.71	1.12	1.16	0.30	0.63	0.89		
$\tau = 0.25$																		
JM	N	PM	1.45	6.68	−2.27	−11.33	2.08	−2.76	−0.05	1.08	−62.57	−0.29	0.11	0.28	0.24	–	65703.4	37.9
		L_{CI}	1.03	4.57	−6.10	−24.92	−3.09	−8.14	−0.13	−13.65	−79.58	−1.03	−1.11	0.07	−0.33	–		
		U_{CI}	1.97	8.30	−0.47	−3.37	5.87	2.04	0.04	15.64	−46.21	0.39	1.13	0.50	0.82	–		
JM	SN	PM	−1.83	9.40	−2.62	−14.33	4.21	−4.07	−0.04	−7.67	−64.61	−0.23	0.13	0.28	0.24	0.86	62723.5	17.9
		L_{CI}	−2.42	6.64	−6.61	−27.99	−0.65	−10.93	−0.11	−21.96	−80.38	−0.76	−0.88	0.08	−0.33	0.82		
		U_{CI}	−1.12	11.32	−0.65	−6.30	8.29	2.51	0.03	6.76	−46.81	0.38	1.11	0.48	0.81	0.88		
$\tau = 0.50$																		
JM	N	PM	10.92	5.48	−3.66	1.91	−0.15	0.23	−0.05	0.57	−61.92	−0.14	0.03	0.30	0.27	–	61876.1	29.8
		L_{CI}	10.59	3.03	−6.21	−3.21	−4.69	−1.65	−0.13	−17.12	−78.06	−0.36	−1.02	0.10	−0.32	–		
		U_{CI}	11.25	7.48	−1.74	5.85	3.55	2.11	0.04	16.04	−44.90	0.06	1.06	0.52	0.91	–		
JM	SN	PM	9.07	7.59	−4.92	−1.57	−1.01	0.60	−0.04	−6.16	−66.39	−0.13	0.03	0.31	0.28	0.85	59069.5	10.3
		L_{CI}	8.59	4.74	−8.23	−5.57	−5.58	−1.94	−0.11	−20.69	−80.97	−0.32	−0.96	0.10	−0.28	0.82		
		U_{CI}	9.54	10.32	−2.23	1.80	3.15	3.10	0.04	8.06	−51.65	0.01	1.04	0.52	0.89	0.88		

Table 1 continued

Method	Model		β_1	β_2	β_3	β_4	β_5	β_6	α_1	α_2	α_3	γ_1	γ_2	γ_3	γ_4	δ	DIC	EPD
$\tau = 0.75$																		
JM	N	PM	20.61	6.58	−2.88	4.53	0.37	−3.16	−0.05	−1.29	−65.22	−0.21	0.03	0.28	0.25	–	66570.1	45.0
		L_{CI}	20.15	4.16	−5.95	−15.73	−4.98	−10.76	−0.14	−17.80	−81.00	−0.73	−0.98	0.08	−0.32	–		
		U_{CI}	21.07	8.42	−0.79	13.56	4.86	2.21	0.04	15.02	−47.54	0.15	1.10	0.49	0.81	–		
JM	SN	PM	19.14	9.64	−2.09	6.09	3.90	−9.54	−0.04	−7.92	−65.39	−0.20	0.19	0.25	0.24	0.85	62192.1	10.8
		L_{CI}	18.65	7.77	−3.63	−13.81	−1.51	−15.50	−0.12	−22.72	−81.48	−0.81	−0.67	0.05	−0.29	0.82		
		U_{CI}	19.62	11.50	−0.69	15.19	7.97	−4.31	0.03	7.22	−48.08	0.51	1.13	0.45	0.82	0.88		
$\tau = 0.95$																		
JM	N	PM	0.25	27.71	−0.77	46.72	8.89	4.07	−0.06	−1.21	−67.67	−0.04	−0.34	0.19	0.22	–	77245.6	143.0
		L_{CI}	−19.08	19.53	−7.37	46.08	6.88	2.45	−0.14	−21.93	−84.58	−1.09	−0.97	0.01	−0.27	–		
		U_{CI}	19.61	37.20	8.19	47.38	10.88	5.69	0.03	14.79	−49.14	0.98	0.39	0.38	0.72	–		
JM	SN	PM	0.02	31.68	−1.84	45.78	11.08	6.04	−0.03	−7.06	−60.83	−0.06	−0.41	0.17	0.22	0.85	71012.9	83.8
		L_{CI}	−20.13	23.38	−8.60	44.96	8.50	3.77	−0.10	−20.90	−79.58	−1.24	−1.35	0.00	−0.24	0.82		
		U_{CI}	19.94	41.62	7.42	46.60	13.56	8.35	0.06	7.16	−44.56	1.08	0.59	0.35	0.72	0.88		

at more extreme quantiles of the viral load. Third, similar to the first-phase viral decay rate, the second-phase viral decay rate is also the smallest at median ($\tau = 0.50$), and becomes larger at more extreme quantiles ($\tau = 0.05$ and $\tau = 0.95$). Overall, the first-phase and the second-phase viral decay rates are varying through the five different quantiles, and the highest are found at quantile $\tau = 0.95$. In comparison of Models N and SN at the same quantile, the estimated results indicate that β_1, β_3, and β_4 (except when $\tau = 0.75$) in Model N are larger than their counterparts in Model SN; in contrast, β_2, and β_5 (except when $\tau = 0.50$) in Model N is smaller than its counterpart in Model SN.

For the parameters in the covariate measurement error model (13) in Model N or Model SN at the five quantiles of $\tau = 0.05, 0.25, 0.50, 0.75$ and 0.95, the estimated results are comparable. In comparison of Model N with Model SN at the same quantile of τ, the estimated α_1 in Model N are slightly smaller than their counterparts in Model SN, while the estimated α_2 and α_3 in Model N is larger than its counterpart in Model SN. Figure 3a shows the difference between population-level estimated CD4 trajectories based on Model N and Model SN at median ($\tau = 0.50$). In addition, it can be seen that as τ increases, the estimated skewness parameter δ with significantly positive value slightly decreases. This finding suggests that there is a significantly positive skewness on the CD4 data and confirms the fact that the distribution of the original CD4 data is right-skewed. Thus, incorporating a skewness parameter in the modeling of the CD4 covariate data is recommended.

In the time-to-death model (14), generally, estimates of all parameters vary through the different quantiles, and Model N and Model SN are comparable at the same quantile. Specifically, from Model SN, the estimated results show that age at seroconversion (γ_3) is significantly associated with risk of death at all quantiles, except $\tau = 0.05$. The strongest association is found at the median ($\tau = 0.50$), and decreases as quantile goes extremely (in both directions) with the weakest and non-significant association at $\tau = 0.05$. This indicates that age at seroconversion is a risk factor of death in general, but not for patients with extremely low viral load. No such significant association and trend is detected for γ_1, γ_2, and γ_4.

To select the best model that fits the data adequately, a Bayesian selection criterion, known as deviance information criterion (DIC) suggested by Spiegelhalter et al. (2002), is used. As with other model selection criteria, we caution that DIC is not intended for identification of the "correct" model, only which one fits the data better. In addition to DIC, a Bayesian approach for checking whether the model fits the data is known as posterior predicting checking. Specifically, expected predictive deviance (EPD), formulated by $EPD = E\{\sum_{i,j}(y_{rep,ij} - y_{obs,ij})^2\}$, where $y_{rep,ij}$ is a replicate of the observed $y_{obs,ij}$, is also evaluated (Gelman et al. 2014). This criterion chooses the model where the discrepancy between predictive values and observed values is the lowest. In order to investigate whether Model SN can provide better fit to the data than Model N, the DIC and EPD values are summarized in Table 1. We found that for the same quantile of τ, both of the DIC and EPD values in Model SN are smaller than their counterparts in Model N. In addition, the smallest DIC is found in Model SN at the median $\tau = 0.50$, while the largest at the quantile of $\tau = 0.05$. This is explained by the smaller amount of information available at more extreme quantiles. In summary, our results suggest that it is very important to assume an SN distribution

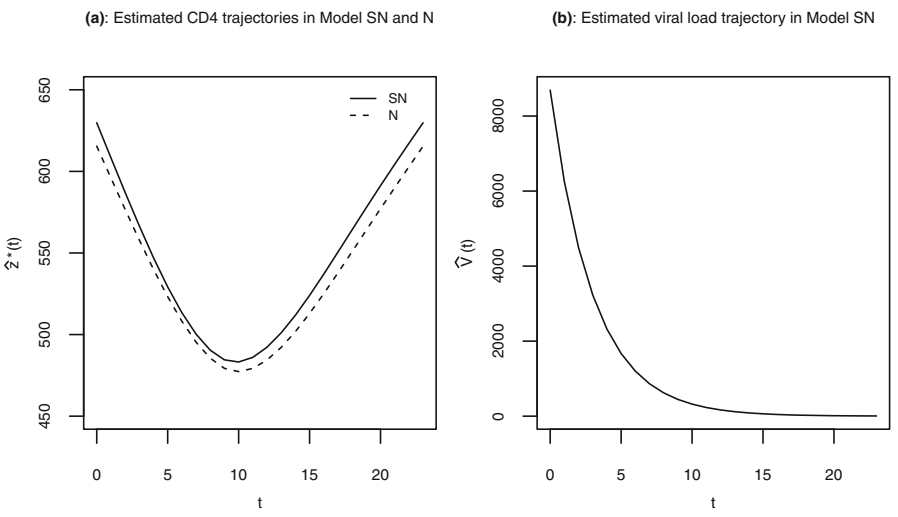

Fig. 3 The population-level estimated CD4 and viral load trajectories at median ($\tau = 0.50$)

for the CD4 covariate model in order to achieve more reliable results, especially when the CD4 data exhibits non-normality.

For a specific application, we further report results of Model SN at the quantile of $\tau = 0.50$. The estimated results indicate that the population level CD4 trajectory may be approximated by the nonparametric mixed-effects model $\hat{z}^*(t) = 325.01(-0.04\psi_0(t) - 6.16\psi_1(t) - 66.39\psi_2(t)) + 556.25$, where $\hat{z}^*(t)$ is in the original CD4 scale (Fig. 3a). The estimated population level first- and second-phase viral decay rates are $\hat{\lambda}_1 = 7.59 - 4.92z_0$ and $\hat{\lambda}_2(t) = -1.01 + 0.60\hat{z}^*(t)$. Thus, the population viral load process may be approximated by $\hat{V}(t) = \exp(9.07 - \hat{\lambda}_1 t) + \exp(-1.57 - \hat{\lambda}_2(t)t)$ (Fig. 3b). Because the first-phase viral decay rate (λ_1) is negatively associated with the baseline CD4 counts, it suggests that the viral load change $V(t)$ may be also significantly associated with the baseline CD4 values. Note that this simple approximation considered here may provide a rough guidance and point to further research even though the true association described above may be complicated. For the covariate measurement error model, the estimate of skewness parameter δ is 0.85 with 95% credible interval: 0.82–0.88, indicating significantly positive skewness. Furthermore, the estimates of parameters in the time-to-death model indicate that higher age at seroconversion will significantly increase the risk of death (Hazard ratio = 1.36, 95% credible interval: 1.11–1.68). Neither of viral load changes over time and race were found to be significantly associated with the risk of death.

5 Simulation studies

To assess the performance of the proposed JM and method, as an illustration, we conduct the following limited simulation studies. The simulated data is similar to the real data used in Sect. 4. Specifically, we generate the data with

sample size n = 300, and assume that each subject had 23 scheduled longitudinal measurements. The simulated measurement time points are similar to those in the real data analysis, and the true parameter values are selected as follows: $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_6)^T = (9, 30, -10, -16, 28, -19)^T$, $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \alpha_3)^T = (5, -15, -70)^T$, $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_4)^T = (1, 1, 5, 5)^T$, $(b_{i1}, \ldots, b_{i4})^T \sim N(\mathbf{0}, \text{diag}(1, 1, 1, 1))$. The time-varying CD4 covariate z_{ij} are generated based on Eq. (13) with $(a_{i1}, a_{i2}, a_{i3})^T \sim N(\mathbf{0}, \text{diag}(1, 1, 1))$; we simulate the model errors e_{ij} and ϵ_{ij} from a $\Gamma(1, 0.5)$ distribution, then subtract by 0.5, which yields a skewed distribution with the mean 0 and variance 2. To generate the survival data, a constant baseline hazard 0.1, and an exponential distribution with mean equal to 0.1 are used to generate censoring time. Other covariates in Cox regression model, race and age, are simulated from a Bernoulli distribution with $p = 0.8$, and a normal distribution with mean 35 and standard deviation 8, respectively. To simulate the left-censoring data, we select the 20% quantile of the longitudinal response as a cut-of threshold, which ensures that 20% of the data is below LOD. According to the settings described above, we only simulate 20 data sets due to exertive computation, and fit the data by Model SN and Model N at three different quantiles of $\tau = 0.25, 0.50$ and 0.75. Note that the prior distributions considered are all non-informative, and same as those in real data analysis. Thus, we expect the results to be robust with respect to prior distributions. Table 2 summarizes simulation results which include the true parameter (TP) values, percent bias (defined by $100 \times \text{bias}_l/|\text{TP}_l|$) and percent mean-square-error (MSE) (defined by $100 \times \sqrt{MSE_l}/|\text{TP}_l|$) of fixed-effects $\boldsymbol{\beta}$, $\boldsymbol{\alpha}$ and $\boldsymbol{\gamma}$.

The magnitude of β_1 and β_4, which indicate that baseline viral load for the first and second phase, respectively, are increasing as the quantiles getting larger. The bias of these two intercept parameters at $\tau = 0.25$ and 0.75 is understandable, because the data is simulated by the mean-regression based model (the estimates of parameters from median regression based models are less biased than those from 25th and 75th QR-based models). For all scenarios considered in this simulation study, it is of interest to see that all estimated biases for β_2 are negative, indicating that these parameters are underestimated, while estimated biases for β_3 and β_6 are positive, suggesting that these parameters are overestimated. Since β_3 is the coefficient of baseline CD4 which is related to the first-phase decay rate and β_6 is the coefficient of time-varying CD4 covariate which is associated with the second-phase decay rate, it might suggest that both of the first-phase decay rate and the second-phase decay rate appear overestimated during the each period of process. However, β_5 are slightly overestimated in Model SN for all quantiles, and underestimated in Model N.

Similarly, for the parameters in the covariate measurement error model, α_2 and α_3 are overestimated, and Model SN outperforms Model N in terms of estimates, bias and MSE, suggesting that it is critical important for covariate model with an SN distribution in order to achieve more accurate and reasonable estimates when the CD4 cell counts exhibit non-normal pattern. This finding may also be explained by the fact that the estimated skewness parameter $\delta = 0.73, 0.72, 0.73$ for $\tau = 0.25, 0.50, 0.75$, respectively, in Model SN. It confirms the significantly right-skewed covariate data. All parameters in the survival component are slightly overestimated in both models, and comparable among different scenarios.

Table 2 Summary of true parameter (TP) values, estimated parameters, Bias and MSE for Models N and SN based on 20 simulated data sets under response model error with $\Gamma(1, 0.5)$ distribution and covariate model error with $\Gamma(1, 0.5)$

TP		Model N			Model SN		
		$\tau = 0.25$	$\tau = 0.50$	$\tau = 0.75$	$\tau = 0.25$	$\tau = 0.50$	$\tau = 0.75$
$\beta_1 = 9$	Bias	−90.73	22.80	135.89	−90.18	21.80	136.02
	MSE	90.74	22.82	135.90	90.19	21.84	136.03
$\beta_2 = 30$	Bias	−28.76	−23.32	−27.64	−28.29	−21.97	−27.64
	MSE	29.38	23.68	28.52	28.66	22.42	27.67
$\beta_3 = -10$	Bias	27.25	23.30	25.67	27.23	20.59	25.62
	MSE	27.43	23.40	26.02	27.36	20.88	25.63
$\beta_4 = -16$	Bias	−61.66	3.34	63.40	−55.32	3.04	67.34
	MSE	61.71	3.51	63.47	55.37	3.94	67.35
$\beta_5 = 28$	Bias	−5.99	−3.49	−6.49	5.75	4.41	1.59
	MSE	6.28	4.22	8.19	6.12	5.16	2.02
$\beta_6 = -19$	Bias	1.97	1.35	1.22	1.10	0.58	1.07
	MSE	2.09	1.43	1.56	1.13	0.71	1.14
$\alpha_1 = 5$	Bias	0.31	0.24	−0.25	−0.89	0.28	2.34
	MSE	1.15	1.38	0.88	2.93	1.47	2.84
$\alpha_2 = -15$	Bias	40.68	39.41	35.01	9.14	8.93	9.15
	MSE	45.01	56.44	47.53	11.60	10.22	11.92
$\alpha_3 = -70$	Bias	7.94	4.47	8.54	0.88	0.63	0.92
	MSE	9.12	4.82	10.07	1.32	1.09	1.17
$\gamma_1 = 1$	Bias	0.41	0.39	0.43	0.40	0.39	0.45
	MSE	0.54	0.48	0.49	0.45	0.49	0.57
$\gamma_2 = 1$	Bias	0.55	0.41	0.64	0.60	0.42	0.68
	MSE	0.72	0.62	0.77	0.76	0.69	0.77
$\gamma_3 = 5$	Bias	0.24	0.23	0.23	0.24	0.24	0.26
	MSE	1.52	1.22	1.47	1.55	1.19	1.50
$\gamma_4 = 5$	Bias	0.34	0.27	0.32	0.31	0.19	0.27
	MSE	1.53	1.32	1.44	1.32	1.10	1.20

EST is average of estimates, Bias and MSE are quantified by percent bias $= 100 \times \text{bias}_l / |\text{TP}_l|$ and percent $\sqrt{\text{MSE}} = 100 \times \sqrt{\text{MSE}_l} / |\text{TP}_l|$, respectively

Overall, the simulation results indicate that median regression-based models (i.e., $\tau = 0.50$) outperform its counterparts (i.e., $\tau = 0.25, 0.75$), while QR-based models with $\tau = 0.25$ and 0.75 are comparable in general. Also, we found that the proposed Model SN provides less biased results compared with Model N. In summary, our simulation results may suggest that it is important to use the QR-NLMET JM where the covariate measurement error model is assumed to have SN distribution, in particular, when the covariate exhibits skewness.

6 Concluding discussion

We presented a Bayesian JM with three components (response, covariate, and survival processes) linked through the random-effects that characterize the underlying individual-specific longitudinal process. In detail, we applied QR-NLMET model with ALD to the longitudinal response (viral load) process, nonparametric LME model with SN distribution to covariate (CD4) processes, and Cox proportional hazard model to the survival process. Although JM for longitudinal–survival data have been an active area of statistical methodological study (Chen et al. 2014; Huang and Chen 2016; Huang et al. 2011), to the best of our knowledge, this is the first time to jointly model response, covariate, and survival processes, with consideration of multiple data features simultaneously, including LOD, covariate measurement error, and skewness. This study is a methodology study and motivated by one real data (MACS), but our method and models are flexible to be generalized and applied to other complex longitudinal–survival data.

Compared to the traditional mean-regression based JM, which only quantify the average, our QR-NLMET JM can fully scan the varying viral load dynamics, and detect the heterogeneous covariates effects at different quantiles. For example, the first-phase and the second-phase viral decay rates are greatly distinct at five different quantiles, where the highest is found at quantile $\tau = 0.95$. It indicates that antiretroviral treatment (ART) is more effective for HIV patients with higher viral load. Interestingly, the first-phase and the second-phase viral decay rates are relatively high at quantile $\tau = 0.05$. This suggests that patients should receive the treatment as early as possible when the viral load is low. Additionally, higher age at seroconversion is associated with the higher risk of death, and the strength of this association varies through different quantiles. Furthermore, we considered SN distribution in covariate measurement error model, instead of normal distribution, when our covariate CD4 exhibits skewness. The Model SN performs better than Model N in both the real data analysis and the simulation studies, and provides more accuracy in parameter estimation. Thus, according to the claim of "evidence-based medicine" and "precision medicine", our QR-NLMET JM has the capability to provide more accurate information/evidence to physicians, benefiting their clinical decision making and treatment evaluation.

For inference of QR-NLMET JM, parameter (or model) identifiability can be an important but difficult problem when many model parameters must be estimated simultaneously. Thus, we prefer parsimonious models that contain fewer parameters, especially for models of secondary interest such as the covariate model whose parameters may be viewed as nuisance parameters. Fortunately, with nonlinear models such as the ones considered in this paper, parameter identifiability is often less of a concern than with linear models (Carroll et al. 2006). In practice, if the models are not identifiable, the MCMC algorithm would diverge quickly. In the application considered in this paper, the MCMC algorithm converged without problems and we did not observe potential identifiability issues.

In Bayesian analysis, sensitivity analysis is important to check the robustness of the posterior estimates when assuming different priors. We performed sensitivity analysis by adopting a few sets of different values for the hyper-parameters of fixed-effects parameters $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, and $\boldsymbol{\gamma}$ in (11). Particularly, independent normal distribution $N(0, 10)$

and $N(5, 100)$ are chosen for each element of them. The close results give us confidence that the posterior estimates are robust to different priors.

The QR-NLMET JM considered in this article can be easily fitted using MCMC procedure via the WinBUGS package interacted with R software that are available publicly. This makes our approach quite powerful and accessible to practical statisticians in the fields. In the real data analysis where $n = 435$, the computing time necessary to fit the proposed model can range up to around 3 days on a Windows PC with Intel Core i7-6700 CPU @ 3.40 GHz and RAM of 16 GB. Due to the intensive computational time for such a complicated modelling approach, we did only 20 replications for each model with different scenarios in the simulation studies, which was relatively small and may have a chance to increase the bias. However, the variation of each parameter of interest was small, indicating that the results of our simulation studies were reliable. This relatively long computing time could likely be lessened by using parallel computing in the future.

Some limitations and issues of this study have to be addressed. First, treatment information is not included in the real data analysis. Lack of this important information may lead to bad model fit. Although this study is a methodology study, we will try to obtain and consider the treatment information in further studies. Second, in simulation studies, the bias of the first phase parameters β_1, β_2, and β_3 is relatively larger than those in the second phase. It can be partially explained by the LOD effect. In our simulated data set, the simulated viral load starts at a low level, then increases as time goes. As a result of this simulated trajectory, the majority of the value below the LOD exists in the first phase. Third, as one of the reviewers pointed out, the MCMC samples are not nearly independent, two options may be able to solve the issue: one is to increase the number of thinning, the other one is to replace the equal-tailed 95% CIs by the dependence-adjusted 95% CIs. To consider this potential issue, we increased the number of total iteration to 200,000 for each chain, and retained every 250th MCMC sample after 100,000 burn-in iterations. We found that both the PMs and the associated 95% CIs (the results not shown here) are very close to the corresponding those presented in Table 1. It is noted that although the posterior summaries are not affected, the 95% CIs may need to be adjusted when the high dependence of the MCMC samples exists.

This article combined new technologies in mathematical modeling and statistical inference with advances in HIV/AIDS dynamics to quantify complex HIV disease mechanism. The complex nature of HIV/AIDS will naturally pose some challenges such as missing values, and multiple failure types. We may further extend our models to more complicated models. For example, (i) consider the missing covariates by introducing simple non-ignorable model for missing mechanism (Huang 2016; Huang and Chen 2016); (ii) jointly model the longitudinal and competing risk processes (Elashoff et al. 2008; Hu et al. 2009). These complicated problems are beyond the focus of this article, but a further study may be warranted. We are actively investigating these interesting issues and hope that we could report these related results in the near future.

Compliance with ethical standards

Conflicts of interest The authors have declared no conflict of interest.

Appendix: Skew-normal distribution and asymmetric Laplace distribution

A.1: Skew-normal distribution

Different versions of multivariate skew distributions have been proposed and used in the literature (Arellano-Valle et al. 2007; Arellano-Valle and Genton 2005; Azzalini and Capitanio 1999; Jara et al. 2008; Sahu et al. 2003). A new class of distributions by introducing skewness in multivariate elliptically distributions, referred as skew-elliptical (SE) distributions, were developed in the literature (Genton 2004; Sahu et al. 2003). The class, which is obtained by using transformation and conditioning, which contains many standard families including the multivariate skew-normal (SN) distribution as special case. A k-dimensional random vector $\boldsymbol{Y}$ follows a k-variate SE distribution if its probability density function (pdf) is given by

$$f(\boldsymbol{y}|\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\Gamma}; m_{\nu}^{(k)}) = 2^k f(\boldsymbol{y}|\boldsymbol{\mu}, \boldsymbol{A}; m_{\nu}^{(k)}) P(\boldsymbol{V} > \boldsymbol{0}), \tag{A.1}$$

where $\boldsymbol{A} = \boldsymbol{\Sigma} + \boldsymbol{\Gamma}^2$, $\boldsymbol{\mu}$ is a location parameter vector, $\boldsymbol{\Sigma}$ is a $k \times k$ positive (diagonal) covariance matrix, $\boldsymbol{\Gamma} = \text{diag}(\delta_1, \delta_2, \ldots, \delta_k)$ is a $k \times k$ skewness matrix with the skewness parameter vector $\boldsymbol{\delta} = (\delta_1, \delta_2, \ldots, \delta_k)^T$; $\boldsymbol{V}$ follows the elliptical distribution $El\left(\boldsymbol{\Gamma}\boldsymbol{A}^{-1}(\boldsymbol{y} - \boldsymbol{\mu}), \boldsymbol{I}_k - \boldsymbol{\Gamma}\boldsymbol{A}^{-1}\boldsymbol{\Gamma}; m_{\nu}^{(k)}\right)$ and the density generator function $m_{\nu}^{(k)}(u) = \frac{\Gamma(k/2)}{\pi^{k/2}} \frac{m_{\nu}(u)}{\int_0^{\infty} r^{k/2-1} m_{\nu}(u) dr}$, with $m_{\nu}(u)$ being a function such that $\int_0^{\infty} r^{k/2-1} m_{\nu}(u) dr$ exists. The function $m_{\nu}(u)$ provides the kernel of the original elliptical density and may depend on the parameter ν. This SE distribution is denoted by $SE(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\Gamma}; m^{(k)})$. One example of $m_{\nu}(u)$, leading to an important special case used throughout the paper, is $m_{\nu}(u) = \exp(-u/2)$. This expression leads to the multivariate SN distribution.

As we know, a normal distribution is a special case of an SN distribution when the skewness parameter is zero. For completeness, this Appendix briefly summarizes the multivariate SN distribution introduced by (Sahu et al. 2003) to be suitable for a Bayesian inference since it is built using the conditional method. For detailed discussions on properties of SN distribution, see publication by (Sahu et al. 2003). Assume a k-dimensional random vector $\boldsymbol{Y}$ follows a k variate SN distribution with location vector $\boldsymbol{\mu}$, $k \times k$ positive (diagonal) covariance matrix $\boldsymbol{\Sigma}$ and $k \times k$ skewness diagonal matrix $\boldsymbol{\Gamma} = \text{diag}(\delta_1, \delta_2, \ldots, \delta_k)$.

A k-dimensional random vector $\boldsymbol{Y}$ follows a k-variate SN distribution, if its pdf is given by

$$f(\boldsymbol{y}|\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\Gamma}) = 2^k |\boldsymbol{A}|^{-1/2} \phi_k\{\boldsymbol{A}^{-1/2}(\boldsymbol{y} - \boldsymbol{\mu})\} P(\boldsymbol{V} > \boldsymbol{0}), \tag{A.2}$$

where $\boldsymbol{V} \sim N_k\{\boldsymbol{\Gamma A}^{-1}(\boldsymbol{y}-\boldsymbol{\mu}), \boldsymbol{I}_k - \boldsymbol{\Gamma A}^{-1}\boldsymbol{\Gamma}\}$, and $\phi_k(\cdot)$ is the pdf of $N_k(\boldsymbol{0}, \boldsymbol{I}_k)$. We denote the above distribution by $SN_k(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\Gamma})$. An appealing feature of Eq. (A.2) is that it gives independent marginal when $\boldsymbol{\Sigma} = diag(\sigma_1^2, \sigma_2^2, \ldots, \sigma_k^2)$. The pdf (A.2) thus simplifies to

$$f(\boldsymbol{y}|\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\Gamma}) = \prod_{i=1}^{k} \left[\frac{2}{\sqrt{\sigma_i^2 + \delta_i^2}} \phi \left\{ \frac{y_i - \mu_i}{\sqrt{\sigma_i^2 + \delta_i^2}} \right\} \Phi \left\{ \frac{\delta_i}{\sigma_i} \frac{y_i - \mu_i}{\sqrt{\sigma_i^2 + \delta_i^2}} \right\} \right], \tag{A.3}$$

where $\phi(\cdot)$ and $\Phi(\cdot)$ are the pdf and cdf of the standard normal distribution, respectively. The mean and covariance matrix are given by

$$E(\boldsymbol{Y}) = \boldsymbol{\mu} + \sqrt{2/\pi}\boldsymbol{\delta}, \quad Cov(\boldsymbol{Y}) = \boldsymbol{\Sigma} + (1 - 2/\pi)\boldsymbol{\Gamma}^2. \tag{A.4}$$

It is noted that when $\boldsymbol{\delta} = \boldsymbol{0}$, the SN distribution reduces to usual normal distribution. In order to have a zero mean vector, we should assume the location parameter $\boldsymbol{\mu} = -\sqrt{2/\pi}\boldsymbol{\delta}$.

According to the study by (Arellano-Valle et al. 2007), if $\boldsymbol{Y}$ follows $SN_k(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\Gamma})$, it can be expressed by a convenient stochastic representation as follows.

$$\boldsymbol{Y} = \boldsymbol{\mu} + \boldsymbol{\Gamma}|\boldsymbol{X}_0| + \boldsymbol{\Sigma}^{1/2}\boldsymbol{X}_1, \tag{A.5}$$

where $\boldsymbol{X}_0$ and $\boldsymbol{X}_1$ are two independent $N_k(\boldsymbol{0}, \boldsymbol{I}_k)$ random vectors. Let $\boldsymbol{w} = |\boldsymbol{X}_0|$; then, $\boldsymbol{w}$ follows a k-dimensional standard normal distribution $N_k(\boldsymbol{0}, \boldsymbol{I}_k)$ truncated in the space $\boldsymbol{w} > \boldsymbol{0}$. Thus, a two-level hierarchical representation of (A.5) is given by

$$\boldsymbol{Y}|\boldsymbol{w} \sim N_k(\boldsymbol{\mu} + \boldsymbol{\Gamma w}, \boldsymbol{\Sigma}), \ \ \boldsymbol{w} \sim N_k(\boldsymbol{0}, \boldsymbol{I}_k)\boldsymbol{I}(\boldsymbol{w} > \boldsymbol{0}). \tag{A.6}$$

Note that when $\boldsymbol{\Gamma} = \boldsymbol{0}$, the hierarchical expression (A.6) presented for the SN distribution $SN_k(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\Gamma})$ reduces to its counterpart for the normal distribution $N_k(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.

A.2: Asymmetric Laplace distribution

An asymmetric distribution, referred as asymmetric Laplace distribution (ALD) which is closely related to the check function for quantile regression (QR), has been discussed in the literature (Geraci and Bottai 2007; Koenker and Machado 1999; Yu and Moyeed 2001; Yu and Zhang 2005). A random variable Y is said to follow ALD if its probability density function (pdf) with parameters μ, σ and τ is given by

$$f(y|\mu, \sigma, \tau) = \frac{\tau(1-\tau)}{\sigma} exp\left\{ -\rho_\tau \left(\frac{y-\mu}{\sigma} \right) \right\}, \tag{A.7}$$

where $\rho_\tau(u) = u(\tau - I(u < 0))$ is the check function, $I(\cdot)$ is the indicator function, $0 < \tau < 1$ is the skewness parameter, $\sigma > 0$ is the scale parameter and $-\infty <$

$\mu < \infty$ is the location parameter. The range of y is $(-\infty, \infty)$. We denote the above distribution by ALD(μ, σ, τ). It should be noted that the check function $\rho_\tau(\cdot)$ assigns weight τ or $1-\tau$ to the observations greater or less than μ, respectively, and that $Pr(y \leq \mu) = \tau$. Therefore, the distribution splits along the scale parameter into two parts, one with probability τ to the left, and one with probability $(1-\tau)$ to the right. That is, ALD(μ, σ, τ) is skewed to left when $\tau > 1/2$, and skewed to right when $\tau < 1/2$. When $\tau = 1/2$, ALD(μ, σ, τ) reduces to the Laplace double exponential (or symmetric Laplace) distribution we usually call which has pdf as follows.

$$f(y|\mu, \sigma, 1/2) = \frac{1}{4\sigma} exp\left\{-\frac{|y-\mu|}{2\sigma}\right\}. \tag{A.8}$$

If $Y \sim$ ALD(μ, σ, τ), then $Pr(y \leq \mu) = \tau$ and $Pr(y > \mu) = 1-\tau$, which shows that the parameters μ and τ in ALD satisfy μ to be the τth quantile of the distribution. This important feature of ALD has been generally adopted for quantile inference (Geraci and Bottai 2007; Yu and Moyeed 2001; Yu et al. 2003) and made it more popular than other asymmetric Laplace distributions (Johnson et al. 1995; Kotz et al. 2002). See (Yu and Zhang 2005) for further properties and generalizations of this distribution. It can be shown that the mean and variance of Y are given by

$$E(Y) = \mu + [\sigma(1-2\tau)]/[\tau(1-\tau)], \quad Var(Y) = [\sigma^2(1-2\tau+2\tau^2)]/[\tau^2(1-\tau)^2]. \tag{A.9}$$

However, the ALD is not smooth and thus difficult to maximize its likelihood function. Fortunately, As shown by (Kotz et al. 2001) and (Kozumi and Kobayashi 2011), the ALD has various mixture representations. To develop Bayesian approach-based sampling algorithms for the QR model, we utilize a hierarchical mixture of exponential and normal distributions (Kotz et al. 2001; Kozumi and Kobayashi 2011). For $Y \sim$ ALD(μ, σ, τ), then Y can be decomposed as the following mixture representation.

$$Y = \mu + \vartheta_1 X_1 + \sqrt{\vartheta_2 \sigma X_1} X_2, \tag{A.10}$$

where X_1 and X_2 are mutually independent, $X_1 \sim Exp(\frac{1}{\sigma})$ with mean σ and $X_2 \sim N(0, 1)$, $\vartheta_1 = (1-2\tau)/[\tau(1-\tau)]$ and $\vartheta_2 = 2/[\tau(1-\tau)]$. This representation can transform the ALD to smooth conditional normal distribution and has been extensively utilized in the recent studies (Kobayashi and Kozumi 2012; Kozumi and Kobayashi 2011; Reich et al. 2012). Thus, a two-level hierarchical representation of (A.10) is given by

$$Y|X_1 \sim N(\mu + \vartheta_1 X_1, \vartheta_2 \sigma X_1), \; X_1 \sim Exp\left(\frac{1}{\sigma}\right). \tag{A.11}$$

A.3: Relationship between nonlinear quantile regression and ALD

Let y_i and $\boldsymbol{x}_i$ denote the outcome of interest and the corresponding covariate vector for subject i $(i = 1, \ldots, n)$, where y_i is independent scalar observations of a continuous

random variable with common cumulative distribution function (cdf) $F_{y_i}(\cdot)$. The τth nonlinear QR model for the response y_i given $\boldsymbol{x}_i$ takes the form of

$$Q_{y_i}(\tau|\boldsymbol{x}_i) = g(\boldsymbol{x}_i, \boldsymbol{\beta}), \tag{A.12}$$

where $Q_{y_i}(\cdot) \equiv F_{y_i}^{-1}(\cdot)$ is the inverse of cdf of y_i given $\boldsymbol{x}_i$ evaluated at τ with $0 < \tau < 1$, $g(\cdot)$ is a nonlinear known function. The nonlinear regression coefficient vector $\boldsymbol{\beta}$ is estimated by minimizing

$$\sum_{i=1}^{n} \rho_\tau \left(y_i - g(\boldsymbol{x}_i, \boldsymbol{\beta})\right), \tag{A.13}$$

where $\rho_\tau(\cdot)$ is the check function defined by $\rho_\tau(u) = u(\tau - I(u < 0))$ and $I(\cdot)$ denotes the indictor function. In order to highlight the τ-distributional dependency, the parameter vector $\boldsymbol{\beta}$ should be indexed by τ (i.e., $\boldsymbol{\beta}(\tau)$). For sake of simplicity, however, we will omit this notation in the reminder of the paper. The check function is closely related to the ALD; see (Koenker and Machado 1999; Yu and Moyeed 2001; Yu and Stander 2007) in detail. The density function of an ALD, denoted by ALD(μ, σ, τ), is briefly discussed in Appendix. Considering σ a nuisance parameter, it can be easily shown that the minimization of Eq. (A.13) with respect to the parameter $\boldsymbol{\beta}$ is exactly equivalent to the maximization of a likelihood function of y_i by assuming y_i from an ALD(μ, σ, τ) with $\mu = g(\cdot)$.

The relationship between the check function and ALD can be used to reformulate the QR method in the likelihood framework. By utilizing this property, under independent data setting, a large number of QR-based statistical models and various associated analysis methods have been investigated in the literature. For example, (Koenker and Machado 1999) proposed a likelihood-based goodness-of-fit test for QR. (Yu and Moyeed 2001) developed Bayesian QR, and (Yu and Stander 2007) and (Kozumi and Kobayashi 2011) studied the Bayesian estimation procedure for the Tobit QR model with censored data. More recently, QR-based linear mixed-effects models have been considered via different methods for longitudinal data (Farcomeni 2012; Geraci and Bottai 2007; Kim and Yang 2012; Koenker 2004; Lipsitz et al. 1997; Liu and Bottai 2009; Wang and Fygenson 2009; Yuan and Yin 2010).

Appendix B: R and WinBUGS program codes for Model SN

```
## R program code
 library(arm)
 library(R2WinBUGS)
 data<-list(......)
 inits<-list()  # Run three chains
 inits[[1]]<-list(......)
 inits[[2]]<-list(......)
 inits[[3]]<-list(......)
```

```
 parameters<-c(......)
 RNAdynamic.QR<-bugs(data, inits, parameters, "Winbugs-
 SN.txt",n.chains=3,
            n.thin=50, n.iter=150000,n.burnin=100000,
bugs.seed=654321,
            bugs.directory="C:\\Users\\WinBUGS14",
DIC=TRUE,debug=TRUE)
## WinBUGS program code named Winbugs-SN.txt
 model
 {
  for (i in 1:n) #total number of subjects
   {
    # Latent random variables generated by SN dist.
    u.e[i] ~ dgamma(d1,d1)
    # Random effects of response model
    b0[i,1]<-0
    b0[i,2]<-0
    b0[i,3]<-0
    b0[i,4]<-0
    b[i,1:4]~dmnorm(b0[i,1:4], Omega[,])
    # Random effects of covariate model
    a3[i,1]<-0
    a3[i,2]<-0
    a3[i,3]<-0
    a[i,1:3]~dmnorm(a3[i,1:3], Omega3[,])
   }
  for (i in 1:n) #Cox regression model
   {for(j in 1:Tnum) {
    risk[i,j] <- step(obs.t[i] - t[j] + eps)
    dN[i,j] <- risk[i, j] * step(t[j + 1] - obs.t[i] -
eps) * fail[i] }
   }
  for(j in 1:Tnum)
   {for(i in 1:n) {
    dN[i, j] ~ dpois(Idt[i, j])
    Idt[i, j] <- risk[i, j]*exp(gamma[1]*b[i,2]+gamma[2]
*b[i,4]+gamma[3]*age[i]+gamma[4]*race[i])*dL0[j] }
    dL0[j] ~ dgamma(mucox[j], c)
    mucox[j] <- dL0.star[j]*c
   }
    c <- 0.001
    r <- 0.1
    for (j in 1 : Tnum)
    { dL0.star[j] <- r * (t[j + 1] - t[j]) }
  for (j in 1:N)  #total number of measurements
   { #Measurement error model with SN distribution
```

```
    z.star[j]<-(alpha[1]+a[y[j,1],1])+(alpha[2]+a[y[j,1],
2])*Z[j,3]+(alpha[3]+a[y[j,1],3])*Z[j,4]
w[j] ~ dnorm(0,1)I(0,)
    z.mu[j]<-z.star[j]+delta*(w[j]-0.798)
    y[j,6] ~ dnorm(z.mu[j], Isigma2)
    # y[j,1]=patid, y[,3]=time, y[j,6]=standardized cd4
    # Response models with ALD dist. (tau=0.50)
    p1[j]       <-beta[1]                          +b[y
[j,1],1]
    lambda1[j]  <-beta[2]+beta[3]*base[y[j,1],3] +b[y
[j,1],2]
    p2[j]       <-beta[4]                          +b[y
[j,1],3]
    lambda2[j]  <-beta[5]+beta[6]*z.star[j]      +b[y
[j,1],4]
    dm1[j]<-p1[j]-step(lambda1[j]-lambda2[j])*lambda1[j]
*y[j,3]
    dm2[j]<-p2[j]-step(lambda1[j]-lambda2[j])*lambda2[j]
*y[j,3]
    dm3[j]<-exp(dm1[j])
    dm4[j]<-exp(dm2[j])
    dm[j]<- dm3[j]+dm4[j]
    v.ald[j] ~ dexp(Isigma)I(1,)
    mu[j]<-log(dm[j])/log(10)+(1-2*tau)/(tau*(1-tau))
*v.ald[j]
    aau[j]<-(tau*(1-tau))/(2*sigma*v.ald[j])
    # LOD
    upper.limit[j] <-Below.detection*y[j,13]+upper.bound
*(1-y[j,13])
    y[j,12]  ~dnorm(mu[j], aau[j])I(,upper.limit[j])
    # y[j,13]=censoring indicator, y[,12]=log(viral load)
   }
    # Prior distributions of the hyperparameters
  #(1) Coefficients
    for(k in 1:6){beta[k]~dnorm(0,0.01)}
    for(k in 1:4){gamma[k]~dnorm(0,0.01)}
    for(k in 1:3){alpha[k]~dnorm(0,0.01)}
  #(2) Skewness parameters
  delta ~ dnorm(0, 0.01)
  #(3) Precision parameters
    Isigma~dgamma(0.01,0.01)
    sigma<-1/Isigma
    Isigma2~dgamma(0.01,0.01)
    sigma2<-1/Isigma2
  #(4) Variance-covariance matrix
    Omega[1:4,1:4] ~ dwish(R[,],4)
```

```
    V1[1:4,1:4]<-inverse(Omega[,])
    Omega3[1:3,1:3] ~ dwish(R3[,],3)
    V3[1:3,1:3]<-inverse(Omega3[,])
   }
##End of model
```

References

Arellano-Valle RB, Genton MG (2005) On fundamental skew distributions. J Multivar Anal 96(1):93–116

Arellano-Valle R, Bolfarine H, Lachos V (2007) Bayesian inference for skew-normal linear mixed models. J Appl Stat 34(6):663–682

Azzalini A, Capitanio A (1999) Statistical applications of the multivariate skew normal distribution. J R Stat Soc 61(3):579–602

Brown ER (2009) Assessing the association between trends in a biomarker and risk of event with an application in pediatric HIV/AIDS. Ann Appl Stat 3(3):1163

Brown ER, Ibrahim JG (2003) A Bayesian semiparametric joint hierarchical model for longitudinal and survival data. Biometrics 59(2):221–228

Brown ER, Ibrahim JG, DeGruttola V (2005) A flexible b-spline model for multiple longitudinal biomarkers and survival. Biometrics 61(1):64–73

Carroll RJ, Ruppert D, Stefanski LA, Crainiceanu CM (2006) Measurement error in nonlinear models: a modern perspective. CRC Press, Boca Raton

Chen Q, May RC, Ibrahim JG, Chu H, Cole SR (2014) Joint modeling of longitudinal and survival data with missing and left-censored time-varying covariates. Stat Med 33(26):4560–4576

Clayton DG (1991) A Monte Carlo method for Bayesian inference in frailty models. Biometrics 47(2):467–485

Dagne GA, Huang Y (2011) Mixed-effects Tobit joint models for longitudinal data with skewness, detection limits, and measurement errors. J Probab Stat 2012:1–19

Dagne G, Huang Y (2012) Bayesian inference for a nonlinear mixed-effects Tobit model with multivariate skew-t distributions: application to AIDS studies. Int J Biostat 8(1)

Davidian M, Giltinan DM (1995) Nonlinear models for repeated measurement data, vol 62. CRC Press, Boca Raton

Davino C, Furno M, Vistocco D (2013) Quantile regression: theory and applications. Wiley, Hoboken

Elashoff RM, Li G, Li N (2008) A joint model for longitudinal measurements and survival data in the presence of multiple failure types. Biometrics 64(3):762–771

Farcomeni A (2012) Quantile regression for longitudinal data based on latent Markov subject-specific parameters. Stat Comput 22(1):141–152

Farcomeni A, Viviani S (2015) Longitudinal quantile regression in the presence of informative dropout through longitudinal–survival joint modeling. Stat Med 34(7):1199–1213

Ganjali M, Baghfalaki T (2015) A copula approach to joint modeling of longitudinal measurements and survival times using monte carlo expectation-maximization with application to aids studies. J Biopharm Stat 25(5):1077–1099

Gelman A, Rubin DB (1992) Inference from iterative simulation using multiple sequences. Stat Sci 7(4):457–472

Gelman A, Carlin JB, Stern HS, Dunson DB, Vehtari A, Rubin DB (2014) Bayesian data analysis, vol 2. CRC Press, Boca Raton

Genton MG (2004) Skew-elliptical distributions and their applications: a journey beyond normality. CRC Press, Boca Raton

Geraci M, Bottai M (2007) Quantile regression for longitudinal data using the asymmetric laplace distribution. Biostatistics 8(1):140–154

He X, Fu B, Fung WK (2003) Median regression for longitudinal data. Stat Med 22(23):3655–3669

Henderson R, Diggle P, Dobson A (2000) Joint modelling of longitudinal measurements and event time data. Biostatistics 1(4):465–480

Hu W, Li G, Li N (2009) A Bayesian approach to joint analysis of longitudinal measurements and competing risks failure time data. Stat Med 28(11):1601–1619

Huang Y (2016) Quantile regression-based Bayesian semiparametric mixed-effects models for longitudinal data with non-normal, missing and mismeasured covariate. J Stat Comput Simul 86(6):1183–1202
Huang Y, Chen J (2016) Bayesian quantile regression-based nonlinear mixed-effects joint models for time-to-event and longitudinal data with multiple features. Stat Med 35(30):5666–5685
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(1):260–269
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, Wu L (2011) Bayesian inference on joint models of HIV dynamics for time-to-event and longitudinal data with skewness and covariate measurement errors. Stat Med 30(24):2930–2946
Jara A, Quintana F, San Martín E (2008) Linear mixed models with skew-elliptical distributions: a Bayesian approach. Comput Stat Data Anal 52(11):5033–5045
Johnson NL, Kotz S, Balakrishnan N (1995) Continuous univariate distribution, vol 2, 2nd edn. Wiley, New York
Kaslow RA, Ostrow DG, Detels R, Phair JP, Polk BF, Rinaldo CR (1987) The Multicenter AIDS Cohort Study: rationale, organization, and selected characteristics of the participants. Am J Epidemiol 126(2):310–318
Kim MO, Yang Y (2012) Semiparametric approach to a random effects quantile regression model. J Am Stat Assoc 106(496):1405–1417
Kobayashi G, Kozumi H (2012) Bayesian analysis of quantile regression for censored dynamic panel data. Comput Stat 27(2):359–380
Koenker R (2004) Quantile regression for longitudinal data. J Multivar Anal 91(1):74–89
Koenker R (2005) Quantile regression, vol 38. Cambridge University Press, Cambridge
Koenker R, Bassett G Jr (1978) Regression quantiles. Econometrica 46:33–50
Koenker R, Machado JA (1999) Goodness of fit and related inference processes for quantile regression. J Am Stat Assoc 94(448):1296–1310
Kotz S, Kozubowski TJ, Podgórski K (2002) Maximum likelihood estimation of asymmetric Laplace parameters. Ann Inst Stat Math 54(4):816–826
Kotz S, Kozubowski TJ, Podgórski K (2001) Asymmetric multivariate Laplace distribution. In: The Laplace distribution and generalizations. Springer, New York, pp 239–272
Kozumi H, Kobayashi G (2011) Gibbs sampling methods for Bayesian quantile regression. J Stat Comput Simul 81(11):1565–1578
Lipsitz SR, Fitzmaurice GM, Molenberghs G, Zhao LP (1997) Quantile regression methods for longitudinal data with drop-outs: application to CD4 cell counts of patients infected with the human immunodeficiency virus. J R Stat Soc 46(4):463–476
Liu Y, Bottai M (2009) Mixed-effects models for conditional quantiles with longitudinal data. Int J Biostat 5(1)
Liu W, Wu L (2007) Simultaneous inference for semiparametric nonlinear mixed-effects models with covariate measurement errors and missing responses. Biometrics 63(2):342–350
Lunn DJ, Thomas A, Best N, Spiegelhalter D (2000) WinBUGS-a Bayesian modelling framework: concepts, structure, and extensibility. Stat Comput 10(4):325–337
Luo Y, Lian H, Tian M (2012) Bayesian quantile regression for longitudinal data models. J Stat Comput Simul 82(11):1635–1649
Perelson AS, Essunger P, Cao Y, Vesanen M, Hurley A, Saksela K, Markowitz M, Ho DD (1997) Decay characteristics of HIV-1-infected compartments during combination therapy. Nature 387:188–191
Reich BJ, Fuentes M, Dunson DB (2012) Bayesian spatial quantile regression. J Am Stat Assoc 106:6–22
Rizopoulos D (2010) Jm: an R package for the joint modelling of longitudinal and time-to-event data. J Stat Softw 35(9):1–33
Rizopoulos D (2011) Dynamic predictions and prospective accuracy in joint models for longitudinal and time-to-event data. Biometrics 67(3):819–829
Rizopoulos D (2012) Joint models for longitudinal and time-to-event data: with applications in R. CRC Press, Boca Raton
Sahu SK, Dey DK, Branco MD (2003) A new class of multivariate skew distributions with applications to Bayesian regression models. Can J Stat 31(2):129–150
Spiegelhalter DJ, Best NG, Carlin BP, Van Der Linde A (2002) Bayesian measures of model complexity and fit. J R Stat Soc 64(4):583–639

Tang AM, Tang NS, Zhu H (2017) Influence analysis for skew-normal semiparametric joint models of multivariate longitudinal and multivariate survival data. Stat Med 36(9):1476–1490
Tian Y, Tian M (2015) Bayesian joint quantile regression for mixed effects models with censoring and errors in covariates. Comput Stat 31(3):1031–1057
Tsiatis AA, Davidian M (2004) Joint modeling of longitudinal and time-to-event data: an overview. Statistica Sinica 14:809–834
Wang HJ, Fygenson M (2009) Inference for censored quantile regression models in longitudinal studies. Ann Stat 37(2):756–781
Wang Y, Taylor JMG (2001) Jointly modeling longitudinal and event time data with application to acquired immunodeficiency syndrome. J Am Stat Assoc 96(455):895–905
Wu L (2002) A joint model for nonlinear mixed-effects models with censoring and covariates measured with error, with application to aids studies. J Am Stat Assoc 97(460):955–964
Wu H, Ding AA (1999) Population HIV-1 dynamics in vivo: applicable models and inferential tools for virological data from AIDS clinical trials. Biometrics 55(2):410–418
Wu H, Zhang JT (2006) Nonparametric regression methods for longitudinal data analysis: mixed-effects modeling approaches, vol 515. Wiley, Hoboken
Wu L, Liu W, Hu X (2010) Joint inference on HIV viral dynamics and immune suppression in presence of measurement errors. Biometrics 66(2):327–335
Yi G, Liu W, Wu L (2011) Simultaneous inference and bias analysis for longitudinal data with covariate measurement error and missing responses. Biometrics 67(1):67–75
Yu K, Moyeed RA (2001) Bayesian quantile regression. Stat Probab Lett 54(4):437–447
Yu K, Stander J (2007) Bayesian analysis of a Tobit quantile regression model. J Econom 137(1):260–276
Yu K, Zhang J (2005) A three-parameter asymmetric Laplace distribution and its extension. Commun Stat Theory Methods 34(9–10):1867–1879
Yu K, Lu Z, Stander J (2003) Quantile regression: applications and current research areas. J R Stat Soc 52(3):331–350
Yuan Y, Yin G (2010) Bayesian quantile regression for longitudinal studies with nonignorable missing data. Biometrics 66(1):105–114