Article

SMMR
STATISTICAL METHODS IN MEDICAL RESEARCH

Statistical Methods in Medical Research
0(0) 1–20

DOI: 10.1177/0962280215586010
smm.sagepub.com
SAGE

Bayesian multivariate augmented Beta rectangular regression models for patient-reported outcomes and survival data

Jue Wang and Sheng Luo

Abstract

Many longitudinal studies (e.g. observational studies and randomized clinical trials) have collected multiple rating scales at each visit in the form of patient-reported outcomes (PROs) in the close unit interval [0,1]. We propose a joint modeling framework to address the issues from the following data features: (1) multiple correlated PROs; (2) the presence of the boundary values of zeros and ones; (3) extreme outliers and heavy tails; (4) the PRO-dependent terminal events such as death and dropout. Our modeling framework consists of a multivariate augmented mixed-effects sub-model based on Beta rectangular distributions for the multiple longitudinal outcomes and a Cox model for the terminal events. The simulation studies suggest that in the presence of outliers, heavy tails, and dependent terminal event, our proposed models provide more accurate parameter estimates than the joint model based on Beta distributions. The proposed models are applied to the motivating Long-term Study-1 (LS-1 study, $n = 1741$) of Parkinson's disease patients.

Keywords

Augmented Beta, Beta regression, Beta rectangular distribution, longitudinal data, Markov chain Monte Carlo, proportional data

1 Introduction

Rating scales are widely utilized for monitoring disease severity and impact. Many longitudinal studies (e.g. observational studies and randomized clinical trials) have collected multiple rating scales at each visit in the form of patient-reported outcomes (PROs), defined as any report of the status of a patient's health condition that comes directly from the patient, without interpretation of the patient's response by a clinician or anyone else.[1] These PRO instruments are usually composed

Department of Biostatistics, The University of Texas Health Science Center at Houston, Houston, TX, USA

Corresponding author:
Sheng Luo, Department of Biostatistics, The University of Texas Health Science Center at Houston, 1200 Pressler Street, Houston, TX 77030, USA.
Email: sheng.t.luo@uth.tmc.edu

of single or multiple items that focus on patient relevant perceptions or experiences, and the sum of the item scores generates a measure of overall impact. For example, EuroQol vertical visual analog scale (EQ-VAS) takes values between 100 (best imaginable health) and 0 (worst imaginable health), on which patients provide a global assessment of their health by placing a mark on a 100-mm visual analog scale. Similarly, the Parkinson's disease questionnaire (PDQ-39) is the rating scale assessing the Parkinson's disease (PD) patients' quality of life (QoL). It is composed of 39 patient-reported items grouped into 8 dimensions including mobility, activities of daily living, emotional well-being, stigma, social support, cognitions, communication, and pain.[2] All items (e.g. "had difficulty walking 100 yards") are scored from 0 (never) to 4 (always), with higher scores representing worse QoL. The sub-scale score for each dimension is obtained by summing the items within each dimension, dividing by the maximum possible score, and multiplying by 100. Therefore, each sub-scale score ranges from 0 to 100. The PDQ-39 summary index (refer to as PDQ-39) is calculated as the arithmetic mean of the sub-scale scores and it also ranges from 0 to 100 with higher scores representing worse QoL. In practice, we divide the EQ-VAS and PDQ-39 values by 100 to rescale to between 0 and 1.

When the outcomes are in the open unit interval $(0, 1)$ (referred to as proportional data as in Kieschnick and McCullough[3]), regression models based on Beta distributions (refer to as Beta regression models)[4] have been increasingly used to directly model the covariate effects on the proportional response through a generalized linear model (GLM) framework. When the outcomes are measured longitudinally, Beta regression models with random effects have been proposed to account for the within-subject correlation.[5,6] However, many observed outcomes can reach the boundary values zero and one which are not included in the support of the Beta distributions. To address this issue, various approaches have been proposed. For example, Galvis et al.[7] proposed a generalized linear mixed model framework by augmenting the probabilities of zeros and ones to the Beta regression model via a zero-and-one-augmented Beta (ZOAB) model. They termed the model as "augmented" model rather than "inflated" model because the Beta distribution does not include zero and one in its support, similar in spirit to Hatfield et al.[8,9] Although the Beta distribution is very flexible, with its pdf taking many shapes (e.g. strictly increasing or decreasing, U-shaped, and unimodal), it considers neither the tail area events nor the outlying events, fails to represent excess variability and over-occurrence of tail-area events,[10,11,12] which could limit its applications for modeling the proportional data. To this end, Hahn[10] proposed a Beta rectangular (BR) distribution, which is a mixture distribution consisting of a Beta distribution and a uniform (rectangular) distribution between 0 and 1. The BR distribution reduces to the Beta distribution when the mixture probability is 0. Comparing with the Beta distribution, the BR distribution assigns more weight to extremal tail-area events, and more probability to the outliers and extremal events. Bayes et al.[12] proposed a BR regression model for cross-sectional data and obtained more robust inference against outlying observations than the Beta regression. However, their model accounts for neither the within-subject correlation in longitudinal data nor the boundary values zeros and ones.

When multiple longitudinal PROs (e.g. EQ-VAS and PDQ-39) measure some aspects of the same disease, significant overlap can be expected to create correlations between PROs of individuals. The correlation between PROs (inter-outcome correlation) needs to be accounted for to provide additional clinical insight by answering questions such as "do PD patients with higher probability of being 1 in the EQ-VAS score (perfect health) are more likely to report high EQ-VAS value or more likely to report 0 in the PDQ-39 score (best QoL) and to report high PDQ-39 value (worse QoL)?". Another issue of the longitudinal studies is that the follow-up process of some individuals may be stopped by some terminal events such as death or dropout. The terminal events are often correlated

with the multiple longitudinal PROs. The presence of such a dependent failure time is often termed as "dependent censoring", ignoring which could lead to biased parameter estimates.[13] To address this issue, joint analyses of the longitudinal outcomes and survival data have been increasingly used.[14,15,16,17,18,19]. In the current context, the joint analyses are essential not only to correct the bias in the parameter estimates of the longitudinal outcomes, but also to answer the following research questions such as "do PD patients with higher probability of being 1 and with higher value in the EQ-VAS score less likely to have a terminal event?" and "do PD patients with higher probability of being 0 and with lower value in the PDQ-39 score less likely to have a terminal event?". Therefore, in this article, we develop a joint modeling framework to address all of the aforementioned issues: (1) multiple correlated PROs; (2) the presence of the boundary values of zeros and ones; (3) extreme outliers and heavy tails; (4) the PRO-dependent terminal events such as death and dropout. The modeling framework consists of a multivariate augmented BR regression sub-model for the multiple longitudinal outcomes confined in the closed unit interval [0, 1] and a Cox proportional hazard sub-model for the dependent censoring event. Two sub-models are linked by shared random effects denoting subject-specific characteristics.

The rest of the article proceeds as follows. In Section 2, we describe a motivating clinical trial, the presence of boundary values, the outlying observations, and the terminal event. In Section 3, we briefly review the Beta and BR distributions and their regression models, develop the augmented BR regression sub-model, and the joint modeling framework. In Section 4, we discuss the Bayesian inference and Bayesian model selection criteria. In Section 5, we conduct an extensive simulation study with three settings to compare the performance of various models. In Section 6, we apply the proposed model to the motivating clinical trial. Concluding remarks and discussions are given in Section 7.

2 A motivating clinical trial

This research is motivated by the Neuroprotection Exploratory Trials in PD Long-term Study-1 (LS-1 study). The LS-1 study is a multicenter, double-blind, phase III study of creatine in patients with early treated PD to assess whether creatine slows PD clinical decline defined by a combination of cognitive, physical, and QoL measures. A total of 1741 patients with early PD were randomly assigned to receive either placebo or creatine. In-person evaluations were conducted at baseline and then annually beginning at 12 months, until the last enrolled participant has completed 5 years of observation. When the LS-1 study was terminated in August 2013 due to the futility of creatine, many participants had extended follow-up (mean 3.9 years, maximum 6.4 years). The LS-1 study represents the largest cohort of patients with early treated PD ever enrolled in a clinical trial and the details of this study can be found in Elm et al.[20] The primary outcomes of interest are EQ-VAS and PDQ-39 scales in this article. During the follow-up, 78 and 323 individuals died and dropped out of the study, respectively. We define a composite endpoint and time to death or dropout as the terminal event. Hence, we observe 401 occurrences of the terminal event in the LS-1 study.

Figure 1 (left panels) displays the histogram of the EQ-VAS and PDQ-39 scores based on all individuals. The presence of a small amount of boundary values (168 occurrences of 100 s in EQ-VAS and 109 occurrences of 0 s in PDQ-39 out of 8227 observations, or 2.04% and 1.32%, respectively), if unaccounted for, is a critical issue for Beta or BR regression models because the boundary values 0 and 1 are out of the supports of Beta and BR distributions. Figure 1 (right panels) displays the longitudinal profiles of the EQ-VAS and PDQ-39 scores of 50 randomly selected individuals. Because PD is a slow progression disease, it is unexpected to observe sudden value change in the outcome variables such as EQ-VAS and PDQ-39, as indicated by the nearly horizontal lowest smooth curve (black solid lines), computed based on all individuals. However, sudden value

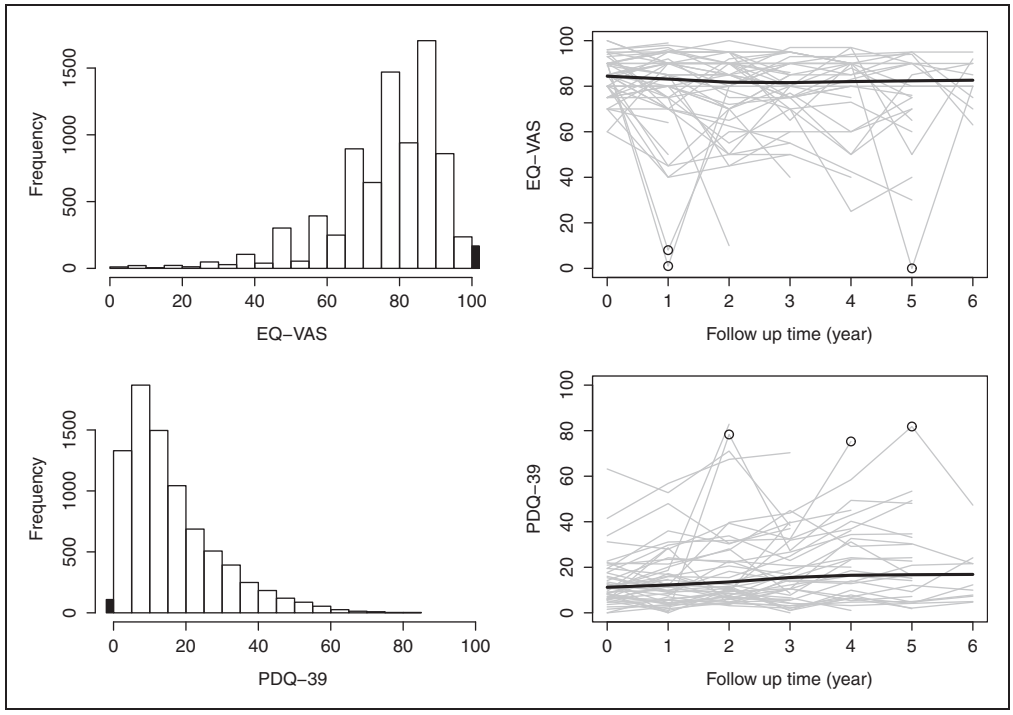

Figure 1. The histograms (left panels) and longitudinal profiles (right panels) of the EQ-VAS (upper panels) and PDQ-39 scores (lower panels) from the LS-1 study and the lowest smooth curve (black solid line). The histograms are based on all individuals while the longitudinal profiles are based on 50 randomly selected individuals. Some observations are circled for further discussion.

changes are observed in both the EQ-VAS and PDQ-39 scores (indicated by the black cycles) in some individuals. Hence, these observations are potential outliers. We divide the original EQ-VAS and PDQ-39 scores by 100 to rescale them as proportional responses confined in the close unit interval $[0, 1]$. We are interested in examining the effect of outliers, as well as the boundary values of 0 and 1, on the inference of regression models based on the Beta and BR distributions, while accounting for the potential dependent terminal event.

3 Model and estimation

3.1 Beta and BR distributions

In this section, we first provide a brief review of the reparameterized Beta distribution[4]. A random variable Y follows a Beta distribution if the probability density function (pdf) in terms of its mean μ and precision parameter ϕ is given by

$$B(Y = y|\mu, \phi) = \frac{\Gamma(\phi)}{\Gamma(\mu\phi)\Gamma((1-\mu)\phi)} y^{\mu\phi-1}(1-y)^{(1-\mu)\phi-1} \quad (1)$$

where $0 < y < 1$, $0 < \mu < 1$, and $\phi > 0$, $E(Y) = \mu$ and $\mathrm{Var}(Y) = \mu(1-\mu)/(1+\phi)$. It is worth mentioning that a natural parameterization of the Beta distribution is via two shape parameters,

i.e. $\alpha = \mu\phi$ and $\beta = (1-\mu)\phi$, and hence $Y \sim \text{Beta}(\alpha,\beta)$. But we adopt the notation $Y \sim \text{Beta}(\mu\phi, (1-\mu)\phi)$ to facilitate the regression analysis on the mean μ. A Beta regression model can be defined under a GLM framework by linking the subject-specific mean μ_i and covariates $\boldsymbol{X}_i$ as $\text{logit}(\mu_i) = \boldsymbol{X}_i\boldsymbol{\beta}$. The precision parameter ϕ can be either constant or log-transformed and then regressed on covariates.

The BR distribution is a mixture distribution consisting of a Beta distribution and a uniform (rectangular) distribution between 0 and 1. Its pdf with support (0, 1) is given by $f(Y = y|\mu,\phi,p) = p + (1-p)B(y|\mu,\phi)$, where $0 \le p \le 1$ is the mixture probability, and $B(y|\mu,\phi)$ is the pdf of the Beta distribution as in (1). The BR distribution is denoted as $Y \sim \text{BR}(\mu,\phi,p)$. If $p = 1$, the BR distribution reduces to the uniform (rectangular) distribution between 0 and 1 and if $p = 0$, it reduces to the Beta distribution $B(y|\mu,\phi)$. The mean and variance of the BR distribution are $E(Y) = (1-p)\mu + \frac{p}{2} = \gamma$ and $\text{Var}(Y) = \frac{\mu(1-\mu)}{1+\phi}(1-p)[1+p(1+\phi)] + \frac{p}{12}(4-3p)$.

To conduct the regression analysis for the mean of the BR distribution, we do the following reparameterization: $\gamma = \frac{p}{2} + (1-p)\mu$ and $\theta = \frac{p}{1-(1-p)|2\mu-1|}$, with $\{0 \le \gamma \le 1,\ 0 \le \theta \le 1\}$.[12] Then, the pdf of the reparameterized BR distribution is

$$f(Y = y|\gamma,\phi,\theta) = \theta(1-|2\gamma-1|) + (1-\theta(1-|2\gamma-1|)) \times B\left(y\middle|\frac{\gamma - 0.5\theta(1-|2\gamma-1|)}{1-\theta(1-|2\gamma-1|)},\phi\right) \quad (2)$$

for $y \in (0, 1)$.

We denote the reparameterized BR distribution (2) as $Y \sim \text{BR}(\gamma,\phi,\theta)$, where γ is the mean, ϕ is the precision parameter, and θ is a shape parameter controlling the thickness of the tails. When the mixture probability $p = 1$, then $\theta = 1$ and $\gamma = 0.5$, the BR distribution reduces to the uniform (rectangular) distribution between 0 and 1. When $p = 0$, then $\theta = 0$ and $\gamma = \mu$, the BR distribution reduces to the Beta distribution. In general, when $0 < p < 1$, then $0 < \theta < 1$, the BR distribution has heavier tails than its Beta distribution counterpart. To visualize this, Figure 2 displays the density functions of various BR distributions with different values of γ, ϕ, and θ. It suggests that when $\theta > 0$, the BR distribution has heavier tails than the corresponding Beta distribution. Similar to the Beta regression model, the BR regression model can be defined as $\text{logit}(\gamma_i) = \boldsymbol{X}_i\boldsymbol{\beta}$ and the precision parameter ϕ can be either constant or regressed on covariates.

3.2 Augmented BR regression model

In this section, we generalize the BR regression model in Section 3.1 to account for the multivariate longitudinal data structure and the boundary values at zero and one. Let $y_{ijk} \in [0, 1]$ be the observed outcome $k (k = 1,\ldots,K$, e.g. outcomes are EQ-VAS, PDQ39 and $K = 2)$ at visit $j (j = 1,\ldots,J_i$, where $j = 1$ is baseline and J_i is the number of visits) from individual $i (i = 1,\ldots,I$, where I is the number of individuals). Let $\mathbf{y}_{ij} = (y_{ij1},\ldots,y_{ijk},\ldots,y_{ijK})'$ be the vector of observation for individual i at visit j, and let $\mathbf{y}_i = (\mathbf{y}_{i1}',\ldots,\mathbf{y}_{iJ_i}')'$ be the outcome vector for individual i across visits, and let $\mathbf{y} = (\mathbf{y}_1',\ldots,\mathbf{y}_I')'$ be the observed outcome vector from all individuals. For outcome k, we propose an augmented BR (ABR) model, denoted by $Y_{ijk} \sim \text{ABR}(p_{0ijk}, p_{1ijk}, \gamma_{ijk}, \phi_k, \theta_k)$, whose pdf follows:

$$f(Y_{ijk} = y_{ijk}|p_{0ijk},p_{1ijk},\gamma_{ijk},\phi_k,\theta_k) = \begin{cases} p_{0ijk} & \text{if } y_{ijk} = 0 \\ p_{1ijk} & \text{if } y_{ijk} = 1 \\ (1-p_{0ijk}-p_{1ijk})f(Y_{ijk} = y_{ijk}|\gamma_{ijk},\phi_k,\theta_k) & \text{if } y_{ijk} \in (0,1) \end{cases} \quad (3)$$

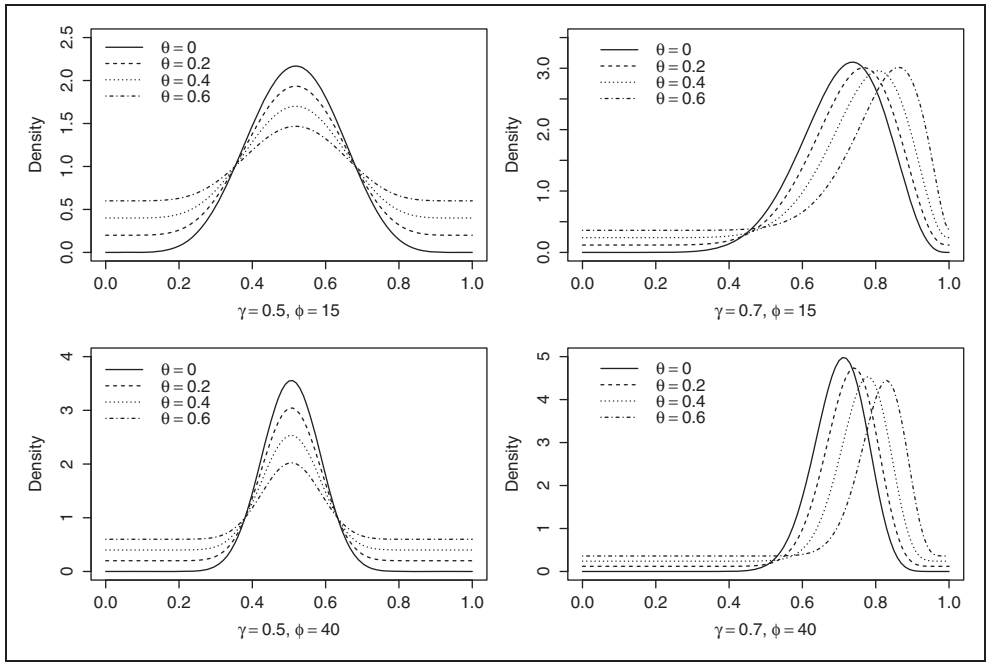

Figure 2. The density functions of various BR distributions with different values of γ, ϕ, and θ. Here $\theta = 0$ (solid line), $\theta = 0.2$ (dashed line), $\theta = 0.4$ (dotted line), $\theta = 0.6$ (dotdash line).

where $p_{0ijk} = P(Y_{ijk} = 0)$, $p_{1ijk} = P(Y_{ijk} = 1)$, $0 \le p_{0ijk} + p_{1ijk} \le 1$, and $f(Y_{ijk} = y_{ijk}|\gamma_{ijk}, \phi_k, \theta_k)$ is the reparameterized BR density function given in (2). Next, we propose the ABR regression model by regressing the covariates onto p_{0ijk}, p_{1ijk}, and the mean γ_{ijk}, which are transformed by appropriate link functions:

$$\begin{aligned} \text{logit}[p_{0ijk} = P(y_{ijk} = 0|u_{i0k})] &= \mathbf{X}_{i0}\boldsymbol{\omega}_{0k} + u_{i0k} \\ \text{logit}[p_{1ijk} = P(y_{ijk} = 1|u_{i1k})] &= \mathbf{X}_{i1}\boldsymbol{\omega}_{1k} + u_{i1k} \\ \text{logit}(\gamma_{ijk}|u_{i2k}) &= \mathbf{X}_{i2}\boldsymbol{\beta}_k + u_{i2k}, \quad \text{for } k = 1, \ldots, K \end{aligned} \tag{4}$$

where the covariate vector $\boldsymbol{X}_{i0}$, $\boldsymbol{X}_{i1}$, and $\boldsymbol{X}_{i2}$ can be identical or different and they include covariates of interest (e.g. treatment assignment) and potential confounding variables (e.g. individual characteristics and socioeconomic status) from individual i. We adopt the logit link function for all three models, while other link functions (e.g. probit and complementary log-log) can also be used. We assume that the random effects vector $\boldsymbol{u}_{ik} = (u_{i0k}, u_{i1k}, u_{i2k})'$ follows a multivariate normal distribution $N(0, \boldsymbol{\Sigma}_k)$, with covariance matrix $\boldsymbol{\Sigma}_k$. When there are a total of K outcomes, the inter-outcome correlation can be modeled by assuming that the random effects vector $\mathbf{u}_i = (\mathbf{u}_1', \ldots, \mathbf{u}_K')'$ follows a multivariate normal distribution with mean 0 and covariance matrix $\boldsymbol{\Sigma}$ (a $3K \times 3K$ matrix).

The proposed ABR regression model can be easily modified to accommodate various features in the data. For example, an augmented Beta (AB) model for outcome k is obtained by replacing the BR density $f(Y_{ijk} = y_{ijk}|\gamma_{ijk}, \phi_k, \theta_k)$ with the Beta density in (1) or equivalently by setting $\theta_k = 0$. If

outcome k only contains zeros or ones, then the ABR regression model (4) can be simplified by removing the parameter p_{0ijk} or p_{1ijk} from model (3). On the other hand, if there are neither zeros nor ones observed in outcome k, we can let $p_{0ijk} = p_{1ijk} \equiv 0$, then the ABR regression model reduces to either the mixed effects BR regression model or the mixed effects Beta regression model (if $\theta_k = 0$). Moreover, additional random effects (e.g. random slopes) can be easily incorporated in model (4).

Let the parameter vector for outcome k be $\boldsymbol{\Theta}_k = (\boldsymbol{\omega}'_{0k}, \boldsymbol{\omega}'_{1k}, \boldsymbol{\beta}'_k, \boldsymbol{\Sigma}'_k, \phi_k, \theta_k)'$ and let the parameter vector for the multiple longitudinal outcomes be $\boldsymbol{\Theta}_y = (\boldsymbol{\Theta}'_1, \ldots, \boldsymbol{\Theta}'_K)$. Conditional on the random effects vector $\boldsymbol{u}_i$, all measurements of each individual are assumed to be independent. The conditional likelihood of the multiple longitudinal outcomes $\boldsymbol{y}_i$ for individual i is

$$\begin{aligned} L_y(\boldsymbol{\Theta}_y; \mathbf{y}_i, \mathbf{u}_i) &= \prod_{k=1}^{K} \prod_{j=1}^{J_i} p(y_{ijk} | \mathbf{u}_{ik}) \\ &= \prod_{k=1}^{K} \prod_{j=1}^{J_i} p_{0ijk}^{I(y_{ijk}=0)} p_{1ijk}^{I(y_{ijk}=1)} \{(1 - p_{0ijk} - p_{1ijk}) f(y_{ijk} | \gamma_{ijk}, \phi_k, \theta_k)\}^{1-I(y_{ijk}=0)-I(y_{ijk}=1)} \end{aligned} \tag{5}$$

where $I(\cdot)$ denotes the indicator function.

Let t_i denote the time to the terminal event for individual i, δ_i (1 if the terminal event is observed and 0 if not) denote the censoring indicator for t_i, and $\boldsymbol{X}_i$ denote the vector of possible risk factors. Vector $\boldsymbol{X}_i$ can share part of or all covariates in vectors $\boldsymbol{X}_{i0}$, $\boldsymbol{X}_{i1}$, and $\boldsymbol{X}_{i2}$. The hazard of having a terminal event at time t_i is

$$\lambda(t_i) = \lambda_0(t_i) \exp(\boldsymbol{X}_i \boldsymbol{\psi} + \sum_{k=1}^{K} \sum_{l=0}^{2} \nu_{kl} u_{ilk}) \tag{6}$$

where $\boldsymbol{\psi}$, $\boldsymbol{\nu} = (\boldsymbol{\nu}'_1, \ldots, \boldsymbol{\nu}'_K)'$ with $\boldsymbol{\nu}'_k = (\nu_{k0}, \nu_{k1}, \nu_{k2})'$ are unknown parameters, and $\lambda_0(t_i)$ is the baseline hazard function. The associated survival function is $S(t_i) = \exp[-\int_0^{t_i} \lambda(s) ds]$. The likelihood of event outcome t_i and δ_i for individual i is $L_s(\boldsymbol{\Phi}_s | t_i, \delta_i, \mathbf{u}_i) = \lambda(t_i)^{\delta_i} S(t_i)$, where the parameter vector $\boldsymbol{\Theta}_s = (\psi', \boldsymbol{\nu}')'$. Conditional on the random effects vector $\boldsymbol{u}_i$, $\boldsymbol{y}_i$ is assumed to be independent of t_i. The full likelihood of the joint model for individual i is

$$L(\boldsymbol{\Theta} | \mathbf{y}_i, t_i, \delta_i, \mathbf{u}_i) = L_y(\boldsymbol{\Theta}_y | \mathbf{y}_i, \mathbf{u}_i) L_s(\boldsymbol{\Theta}_s | t_i, \delta_i, \mathbf{u}_i) \, p(\mathbf{u}_i) \tag{7}$$

where $p(\boldsymbol{u}_i)$ is the density function of $\boldsymbol{u}_i$, the parameter vector $\boldsymbol{\Theta} = (\boldsymbol{\Theta}'_y, \boldsymbol{\Theta}'_s)'$. The "cross-equation correlation" between models (4) and (6) is introduced by the subject level random effects vector $\boldsymbol{u}_i$. Specifically, for the outcome PDQ-39, negative parameter ν_{k0} indicates that the individuals with higher probabilities of being at 0 (good QoL) are less likely to die or drop out of the study. Similarly, positive parameter ν_{k2} suggests that the individuals with lower PDQ-39 values (better QoL) tend to have lower risk of death or dropout. Similar interpretation can be made to the parameters ν_{k1} and ν_{k2} for the outcome EQ-VAS. We refer to the proposed joint modeling framework assuming the BR and Beta distributions as models $\mathrm{JM_{BR}}$ and $\mathrm{JM_{Be}}$, respectively. In order to illustrate the need of joint modeling the longitudinal measurements and survival events, we consider the reduced models which assuming that the survival time is independent of the longitudinal outcomes. We refer to the reduced models assuming the BR and Beta distributions as models $\mathrm{RM_{BR}}$ and $\mathrm{RM_{Be}}$, respectively.

4 Bayesian inference

4.1 Prior specification

The full likelihood formulation in model (7) involves unspecified baseline hazard function $\lambda_0(\cdot)$. We adopt a piecewise constant function to approximate $\lambda_0(\cdot)$. It has been shown that survival models using a piecewise constant baseline hazard function yield good estimators for both fixed effects and frailty,[21,22] although fixed cut points need to be specified a priori. It is more flexible than the a priori choices of baseline hazard distribution (e.g. Weibull distribution) and it retains enough model structure.[23] Specifically, we divide the follow-up period into m intervals by every $1/m$th quantiles denoted by $\tau_1, \tau_2, \ldots, \tau_m$ and $\tau_0 = 0$ or the smallest failure time. The baseline hazard vector is $\boldsymbol{g} = (g_1, g_2, \ldots, g_m)$, and the piecewise constant baseline hazard function is defined as $\lambda_0(t) = \sum_{l=1}^{m} g_l I_l(t)$, with indicator function $I_l(t) = 1$ if $\tau_{l-1} \leq t < \tau_l$ and 0 otherwise.

To make inference on the unknown parameter vector $\boldsymbol{\Theta}$, we use Bayesian inference based on Markov chain Monte Carlo (MCMC) posterior simulations. We use vague priors on all elements in the parameter vector $\boldsymbol{\Theta}$. Specifically, the prior distributions of all elements in the regression parameter vectors $\boldsymbol{\omega}$ and $\boldsymbol{\beta}$ are $N(0, 100)$, where $\boldsymbol{\omega} = (\boldsymbol{\omega}'_1, \ldots, \boldsymbol{\omega}'_K)'$, $\boldsymbol{\omega}'_k = (\boldsymbol{\omega}'_{0k}, \boldsymbol{\omega}'_{1k})'$, and $\boldsymbol{\beta} = (\boldsymbol{\beta}'_1, \ldots, \boldsymbol{\beta}'_K)'$. We use the prior distribution Gamma(0.001, 0.001) for the precision parameter ϕ_k and all of the variances in the covariance matrix $\boldsymbol{\Sigma}$. We use Uniform(-1, 1) prior distribution for all of the correlation coefficients in matrix $\boldsymbol{\Sigma}$ and Uniform(0, 1) prior distribution for the shape parameter θ_k. We have investigated other selections of vague prior distributions and have obtained very similar results in both the simulation studies and the application.

The posterior samples are obtained from the full conditional of each unknown parameter using Hamiltonian Monte Carlo (HMC)[24,25] and No-U-Turn Sampler (NUTS).[26] The HMC and NUTS samplers are implemented in `Stan`, which is a probabilistic programming language implementing statistical inference with HMC and NUTS samplers. The model fitting is performed in `STAN` (version 2.5.0)[27] by specifying the full likelihood function and the prior distributions of all unknown parameters. For large datasets, `STAN` may be more efficient than `BUGS` in achieving faster convergence and requiring smaller number of samples.[26] To facilitate easy reading and implementation of the proposed models, a sample `STAN` code for fitting model $\mathrm{JM_{BR}}$ has been posted in the online supplement (available at: http://smm.sagepub.com/).

To monitor Markov chain convergence, we use history plots and view the absence of apparent trend in the plots as evidence of convergence. We run multiple chains with diffuse initial values and ensure the scale reduction $\hat{R}$ of all parameters are smaller than 1.1.[28]

4.2 Bayesian model selection and influence diagnostics

There are a wide variety of model selection criteria in Bayesian inference. Because of the mixture framework in our model, we use the $\mathrm{DIC_3}$ measurement.[29] The $\mathrm{DIC_3}$ is defined as $DIC_3 = \overline{D(\boldsymbol{\Theta})} + \tau_D$, where $\overline{D(\boldsymbol{\Theta})} = -2\mathrm{E}_{\boldsymbol{\Theta}|\mathbf{D}}\{\log[\prod_{i=1}^{I} f(\mathbf{y}_i, t_i|\boldsymbol{\Theta})]\}$ is the posterior mean deviance, $\tau_D = \overline{D(\boldsymbol{\Theta})} + 2\log\{\mathrm{E}_{\boldsymbol{\Theta}|\mathbf{D}}[\prod_{i=1}^{I} f(\mathbf{y}_i, t_i|\boldsymbol{\Theta})]\}$ is a measure of the effective number of parameters in the model, and $\mathrm{E}_{\boldsymbol{\Theta}|\mathbf{D}}(.)$ is the expectation with respect to the joint posterior distribution $\pi(\boldsymbol{\Theta}|\mathbf{D})$, $\boldsymbol{D}$ denoting the observed data. Thus,

we have $DIC_3 = -4\mathrm{E}_{\boldsymbol{\Theta}|\mathbf{D}}\{\log[\prod_{i=1}^{I} f(\mathbf{y}_i, t_i|\boldsymbol{\Theta})]\} + 2\log\{\mathrm{E}_{\boldsymbol{\Theta}|\mathbf{D}}[\prod_{i=1}^{I} f(\mathbf{y}_i, t_i|\boldsymbol{\Theta})]\}$. Applying Monte Carlo approximation,

$$\widehat{DIC_3} = -\frac{4}{M}\sum_{m=1}^{M}\sum_{i=1}^{I}\log\{f(\mathbf{y}_i, t_i|\boldsymbol{\Theta}^{(m)})\} + 2\log\{\frac{1}{M}\sum_{m=1}^{M}\prod_{i=1}^{I} f(\mathbf{y}_i, t_i|\boldsymbol{\Theta}^{(m)})\},$$

where $\boldsymbol{\Theta}^{(m)}$ is the mth ($m = 1, \ldots, M$) post burn-in posterior samples of parameter vector $\boldsymbol{\Theta}$. A smaller value of DIC_3 indicates a better-fitting model. In addition, we use the expected Akaike information criterion (EAIC) and the expected Bayesian (or Schwarz) information criterion (EBIC) as model selection tools.[30] The EAIC and EBIC can be estimated as $\mathrm{EAIC} = \overline{D(\boldsymbol{\Theta})} + 2p$ and $\mathrm{EBIC} = \overline{D(\boldsymbol{\Theta})} + p\log(I)$, where p is the number of parameters in the parameter vector $\boldsymbol{\Theta}$, and I is the number of individuals. Smaller values of EAIC and EBIC indicate better predictive ability of the model.

Moreover, Bayes factors (BF) is a Bayesian alternative to p values for testing hypotheses and for quantifying the degree to which observed data support or conflict with a hypothesis. Let two competing models be M_1 and M_2. The BF in favor of model M_1 over M_2 is defined as

$$\mathrm{BF}_{12} = \mathrm{BF}(M_1; M_2) = \frac{p(M_1|\mathbf{y})/p(M_2|\mathbf{y})}{p(M_1)/p(M_2)} = \frac{p(\mathbf{y}|M_1)}{p(\mathbf{y}|M_2)}$$

where $p(M_l)$ is the prior probability of model M_l, where $l = 1, 2$, $p(M_l|\mathbf{y})$ is the posterior probability of model M_l, and $p(\mathbf{y}|M_l)$ is the predictive probability of observing data $\boldsymbol{y}$ under model M_l, and $p(\mathbf{y}|M_l) = \int f(\mathbf{y}|\boldsymbol{\Theta}_l, M_l)\, p(\boldsymbol{\Theta}_l|M_l) d\boldsymbol{\Theta}_l$, where $p(\boldsymbol{\Theta}_l|M_l)$ is the prior distribution for parameter vector $\boldsymbol{\Theta}_l$ under model M_l. When $\mathrm{BF}_{12} > 150$, decisive evidence is shown in favor of model M_1; when BF_{12} is between 20 and 150, strong evidence is shown in favor of model M_1.[31] To avoid the integral involved in computation of BF, the Laplace–Metropolis estimator based on the normal distribution[32] is adopted to approximate the predictive probability. Specifically, $p(\mathbf{y}|M_l) \approx (2\pi)^{d_l/2}|\boldsymbol{\Sigma}_l|^{1/2} f(\mathbf{y}|\overline{\boldsymbol{\Theta}}_l, M_l)\, p(\overline{\boldsymbol{\Theta}}_l|M_l)$, where d_l is the number of the parameters in parameter vector $\boldsymbol{\Theta}_l$, $\boldsymbol{\Sigma}_l$ is the posterior covariance matrix of $\boldsymbol{\Theta}_l$, $\overline{\boldsymbol{\Theta}}_l$ is the posterior mean of $\boldsymbol{\Theta}_l$, $p(\overline{\boldsymbol{\Theta}}_l|M_l)$ is the prior probability of parameters evaluated at $\overline{\boldsymbol{\Theta}}_l$, and $f(\mathbf{y}|\overline{\boldsymbol{\Theta}}_l, M_l)$ is the likelihood from model M_l when parameters are at the posterior mean values.

Moreover, the widely used criteria conditional predictive ordinate (CPO)[30,33,34,35,36] is adopted to assess model fit and selection. The CPO for the (ijk)th observation is defined as $\mathrm{CPO}_{ijk} = p(y_{ijk}|y_{(ijk)}) = \int p(y_{ijk}|\boldsymbol{\Theta})\, p(\boldsymbol{\Theta}|y_{(ijk)}) d\boldsymbol{\Theta}$, where y_{ijk} denotes the full data and $y_{(ijk)}$ denotes the data after deleting the (ijk)th observation. CPO is a form of cross-validation with high value indicating that the data for observation (ijk) can be accurately predicted by a model based on the data from all other observations. Hence, a model with larger CPO_{ijk} for all observations suggests a better fit. Although the closed form of CPO_{ijk} is not available for our proposed model, a Monte Carlo estimator of CPO_{ijk} can be obtained by MCMC samples $\{\boldsymbol{\Theta}^{(t)}\}_{t=1}^{M}$ from posterior distribution $p(\boldsymbol{\Theta}|\mathbf{y})$, with M being the total number of post burn-in samples. Because $p(y_{ijk}|y_{(ijk)}) = p(\mathbf{y})/p(y_{(ijk)}) = 1/\int p(\boldsymbol{\Theta}|\mathbf{y})/p(y_{ijk}|y_{(ijk)}, \boldsymbol{\Theta}) d\boldsymbol{\Theta}$, a harmonic-mean approximation of CPO_{ijk} is $\widehat{\mathrm{CPO}}_{ijk} = (\frac{1}{M}\sum_{t=1}^{M}\frac{1}{p(y_{ijk}|y_{(ijk)}, \boldsymbol{\Theta}^{(t)})})^{-1} = (\frac{1}{M}\sum_{t=1}^{M}\frac{1}{p(y_{ijk}|\boldsymbol{\Theta}^{(t)})})^{-1}$.[34] A summary statistics of $\widehat{\mathrm{CPO}}_{ijk}$ for all individuals is the log pseudo-marginal

likelihood (LPML) defined as $\text{LPML} = \sum_{i=1}^{I}\sum_{j=1}^{J_i}\sum_{k=1}^{K}\log(\widehat{\text{CPO}}_{ijk})$. A larger value of LPML indicates better fit of the model.

To detect the occurrence of outliers and extremal events around the tail-area, we consider the Kullback–Leibler (K-L) divergence defined as $K\{p(\boldsymbol{\Theta}|\mathbf{y}), p(\boldsymbol{\Theta}|\, y_{(ijk)})\} = \int p(\boldsymbol{\Theta}|\mathbf{y})\log\left(\frac{p(\boldsymbol{\Theta}|\mathbf{y})}{p(\boldsymbol{\Theta}|\, y_{(ijk)})}\right)d\boldsymbol{\Theta}$. Peng and Day[37] pointed out that $K\{p(\boldsymbol{\Theta}|\mathbf{y}), p(\boldsymbol{\Theta}|\, y_{(ijk)})\} = \log E_{\boldsymbol{\Theta}|\mathbf{y}}[\{p(\, y_{ijk}|\boldsymbol{\Theta})\}^{-1}] + E_{\boldsymbol{\Theta}|\mathbf{y}}[\log\{p(\, y_{ijk}|\boldsymbol{\Theta})\}] = -\log(\text{CPO}_{ijk}) + E_{\boldsymbol{\Theta}|\mathbf{y}}[\log\{p(\, y_{ijk}|\boldsymbol{\Theta})\}]$, where $E_{\boldsymbol{\Theta}|\mathbf{y}}(\cdot)$ denotes the expectation with respect to the joint posterior distribution $p(\boldsymbol{\Theta}|\mathbf{y})$. Cancho et al.[38] proposed the Monte Carlo estimate of the K-L divergence as $\widehat{K}\{p(\boldsymbol{\Theta}|\mathbf{y}), p(\boldsymbol{\Theta}|y_{(ijk)})\} = -\log(\widehat{\text{CPO}}_{ijk}) + \frac{1}{M}\sum_{t=1}^{M}\log\{p(y_{ijk}|\boldsymbol{\Theta}^{(t)})\}$. Hence, we adopt the Monte Carlo estimate to compute the K-L divergence.

5 Simulation studies

In this section, we conduct an extensive simulation study with three settings to compare the performance of the proposed joint models JM_{BR} and JM_{Be}. The simulated data structure is similar to the motivating LS-1 study. Specifically, in all three settings, we generate 200 datasets with sample size $N = 600$ individuals and seven visits, i.e., baseline and six follow-up visits, $J_i = 7$ with the time vector $\boldsymbol{t}_i = (t_{i1}, t_{i2}, \ldots, t_{i7})' = (0, 1, 2, 3, 4, 5, 6)'$. We generate two continuous proportional outcome ($K = 2$) restricted in the interval (0, 1] (assuming that both outcomes can reach boundary value at 1, but not at 0). We consider one covariate x_i taking value 0 or 1 each with probability 1/2 to mimic the treatment assignment. In all settings, the datasets are simulated from the following models:

$$\begin{aligned}
&\text{logit}[P(\, y_{ijk} = 1|u_{i1k})] = \mathbf{X}_{i1}\boldsymbol{\omega}_k + u_{i1k}\\
&\text{logit}(\gamma_{ijk}|u_{i2k}) = \mathbf{X}_{i2}\boldsymbol{\beta}_k + u_{i2k}, \quad \text{for } k = 1, 2\\
&\lambda(t_i) = \lambda_0(t_i)\exp[\mathbf{X}_i\boldsymbol{\psi} + \sum_{k=1}^{2}(\nu_{k1}u_{i1k} + \nu_{k2}u_{i2k})]
\end{aligned}$$

The covariate vectors are $\mathbf{X}_{i1} = (1, x_i)'$, $\mathbf{X}_{i2} = (1, x_i, t_{ij}, x_i t_{ij})'$, and $\boldsymbol{X}_i = x_i$. We set the regression coefficient vectors to be $\boldsymbol{\omega}_1 = (-1.5, -0.5)'$, $\boldsymbol{\omega}_2 = (-1, -1)'$, $\boldsymbol{\beta}_1 = (1.5, -0.5, -0.1, 0.2)'$, $\boldsymbol{\beta}_2 = (1, -1, -0.2, 0.5)'$, and $\psi = -1$. We set the precision parameters to be $\phi_1 = 10$ and $\phi_2 = 5$. The random effects vector $\boldsymbol{u}_i = (u_{i11}, u_{i21}, u_{i12}, u_{i22})'$ is simulated from a multivariate normal distribution with mean 0 and covariance matrix $\boldsymbol{\Sigma}$ generated with the following parameters: $\sigma_{11}^2 = 1.44$, $\sigma_{22}^2 = 0.36$, $\sigma_{33}^2 = 0.64$, $\sigma_{44}^2 = 0.25$, $\rho_{12} = 0.4$, $\rho_{13} = 0.2$, $\rho_{14} = 0.5$, $\rho_{23} = 0.1$, $\rho_{24} = 0.2$, and $\rho_{34} = 0.2$. In all three settings, we assume that the two longitudinal outcomes are associated with the survival outcome with the parameters $\boldsymbol{\nu} = (\boldsymbol{\nu}'_1, \boldsymbol{\nu}'_2)' = (0.2, 0.4, 0.5, 0.4)'$.

We apply the Bayesian framework in Section 4 to obtain inference. To determine the burn-in iterations and assess the MCMC convergence and mixing properties, we examine the trace plots and the autocorrelations. We find that the chains converge reasonably fast and all achieve stationarity within 2000 iterations with $\hat{R}$ of all parameters smaller than 1.1. We run two parallel MCMC chains with diverse initial values and choose 2000 iterations for burn-in and the inference is based on the subsequent 2000 iterations from both chains.

Table 1. Setting I: simulation results from models JM_{Be} and JM_{BR} when data are simulated from model JMBe.

	JM_{Be}				JM_{BR}			
	Bias	SD	CP	RMSE	Bias	SD	CP	RMSE
For the first longitudinal outcome								
$\omega_{10} = -1.5$	0.014	0.116	0.940	0.116	0.010	0.117	0.925	0.117
$\omega_{11} = -0.5$	−0.014	0.160	0.930	0.160	−0.007	0.161	0.920	0.161
$\beta_{10} = 1.5$	−0.005	0.055	0.945	0.055	−0.004	0.055	0.940	0.055
$\beta_{11} = -0.5$	0.001	0.070	0.935	0.070	−0.000	0.071	0.935	0.071
$\beta_{12} = -0.1$	0.000	0.013	0.940	0.013	−0.000	0.013	0.945	0.013
$\beta_{13} = 0.2$	0.001	0.016	0.935	0.016	0.002	0.017	0.930	0.017
$\phi_1 = 10$	−0.021	0.308	0.955	0.308	0.070	0.310	0.960	0.317
$\theta_1 = 0$					0.005	0.002		0.005
For the second longitudinal outcome								
$\omega_{20} = -1$	0.001	0.083	0.940	0.083	0.003	0.081	0.950	0.081
$\omega_{21} = -1$	−0.000	0.130	0.935	0.130	−0.003	0.127	0.945	0.127
$\beta_{20} = 1$	0.000	0.055	0.970	0.055	0.002	0.054	0.965	0.054
$\beta_{21} = -1$	−0.001	0.072	0.955	0.072	−0.000	0.072	0.950	0.072
$\beta_{22} = -0.2$	−0.001	0.015	0.960	0.015	−0.001	0.015	0.965	0.015
$\beta_{23} = 0.5$	0.001	0.021	0.940	0.021	0.001	0.020	0.955	0.020
$\phi_2 = 5$	−0.015	0.152	0.955	0.152	0.081	0.167	0.945	0.186
$\theta_2 = 0$					0.016	0.006		0.017
For the survival outcome								
$\psi = -1$	−0.050	0.166	0.940	0.173	−0.044	0.165	0.950	0.171
$\nu_{11} = 0.2$	0.005	0.181	0.980	0.181	−0.001	0.190	0.965	0.190
$\nu_{12} = 0.4$	0.021	0.241	0.965	0.241	0.014	0.253	0.950	0.253
$\nu_{21} = 0.5$	0.129	0.284	0.945	0.311	0.135	0.268	0.930	0.299
$\nu_{22} = 0.4$	−0.002	0.388	0.965	0.387	0.015	0.412	0.960	0.411
For the random effects covariance matrix								
$\sigma_{11} = 1.44$	−0.004	0.263	0.910	0.263	0.007	0.263	0.925	0.263
$\sigma_{12} = 0.288$	−0.011	0.067	0.925	0.068	−0.007	0.067	0.930	0.068
$\sigma_{13} = 0.192$	−0.014	0.097	0.970	0.098	−0.010	0.097	0.980	0.097
$\sigma_{14} = 0.3$	−0.008	0.053	0.925	0.054	−0.006	0.056	0.920	0.056
$\sigma_{22} = 0.36$	−0.001	0.032	0.950	0.032	0.001	0.034	0.945	0.034
$\sigma_{23} = 0.048$	−0.000	0.048	0.900	0.048	0.002	0.047	0.905	0.047
$\sigma_{24} = 0.06$	−0.001	0.019	0.950	0.019	−0.002	0.020	0.940	0.020
$\sigma_{33} = 0.64$	−0.022	0.143	0.945	0.145	−0.014	0.141	0.965	0.141
$\sigma_{34} = 0.08$	−0.001	0.043	0.950	0.043	0.001	0.041	0.965	0.041
$\sigma_{44} = 0.25$	−0.001	0.029	0.935	0.029	−0.001	0.030	0.925	0.030

In simulation setting I, we simulate data from model JM_{Be} assuming no outliers and extremal events around the tail-areas in both longitudinal outcomes ($\theta_1 = \theta_2 = 0$). Table 1 displays bias (the average of the posterior means minus the true values), standard deviation (SD, the standard deviation of the posterior means), coverage probabilities (CP) of 95% equal-tail credible intervals, and root mean squared error (RMSE) from models JM_{Be} and JM_{BR}. The results suggest that when data come from model JM_{Be}, both models JM_{Be} and JM_{BR} generate comparable results with very small bias and RMSE, and the coverage probability being

Table 2. Setting II: simulation results from models JM_{Be} and JM_{BR} when data are simulated from model JMBR.

	JM_{Be}				JM_{BR}			
	Bias	SD	CP	RMSE	Bias	SD	CP	RMSE
For the first longitudinal outcome								
$\omega_{10} = -1.5$	0.009	0.112	0.930	0.112	0.008	0.109	0.940	0.109
$\omega_{11} = -0.5$	−0.007	0.153	0.955	0.153	−0.002	0.149	0.965	0.148
$\beta_{10} = 1.5$	−0.092	0.062	0.650	0.111	−0.003	0.059	0.935	0.059
$\beta_{11} = -0.5$	0.031	0.076	0.930	0.082	0.000	0.077	0.935	0.076
$\beta_{12} = -0.1$	0.006	0.015	0.915	0.016	−0.000	0.013	0.955	0.013
$\beta_{13} = 0.2$	−0.012	0.020	0.905	0.023	0.001	0.017	0.945	0.017
$\phi_1 = 10$	−4.503	0.306	0.000	4.514	0.021	0.502	0.935	0.502
$\theta_1 = 0.2$					0.002	0.023	0.940	0.023
For the second longitudinal outcome								
$\omega_{20} = -1$	−0.006	0.076	0.965	0.076	−0.004	0.078	0.960	0.078
$\omega_{21} = -1$	0.008	0.117	0.960	0.117	0.008	0.115	0.960	0.115
$\beta_{20} = 1$	−0.034	0.059	0.910	0.068	−0.006	0.059	0.925	0.059
$\beta_{21} = -1$	0.035	0.080	0.880	0.087	0.006	0.077	0.930	0.077
$\beta_{22} = -0.2$	0.007	0.017	0.905	0.019	0.002	0.017	0.950	0.017
$\beta_{23} = 0.5$	−0.015	0.023	0.890	0.027	−0.003	0.022	0.950	0.023
$\phi_2 = 5$	−1.014	0.146	0.000	1.025	0.009	0.248	0.975	0.248
$\theta_2 = 0.15$					0.001	0.032	0.970	0.032
For the survival outcome								
$\psi = -1$	−0.015	0.171	0.960	0.171	−0.009	0.170	0.965	0.170
$\nu_{11} = 0.2$	0.008	0.197	0.930	0.197	0.017	0.193	0.950	0.193
$\nu_{12} = 0.4$	0.067	0.323	0.935	0.329	−0.006	0.259	0.955	0.258
$\nu_{21} = 0.5$	0.098	0.259	0.950	0.276	0.094	0.255	0.945	0.272
$\nu_{22} = 0.4$	0.023	0.460	0.930	0.459	0.003	0.437	0.925	0.435
For the random effects covariance matrix								
$\sigma_{11} = 1.44$	0.011	0.252	0.935	0.251	0.014	0.252	0.955	0.252
$\sigma_{12} = 0.288$	−0.019	0.067	0.890	0.069	−0.007	0.066	0.930	0.066
$\sigma_{13} = 0.192$	−0.002	0.106	0.965	0.106	−0.002	0.103	0.970	0.102
$\sigma_{14} = 0.3$	−0.015	0.051	0.955	0.053	−0.007	0.053	0.965	0.053
$\sigma_{22} = 0.36$	−0.048	0.035	0.670	0.059	0.000	0.035	0.920	0.035
$\sigma_{23} = 0.048$	0.003	0.042	0.950	0.042	0.005	0.040	0.965	0.041
$\sigma_{24} = 0.06$	−0.005	0.020	0.945	0.020	−0.000	0.019	0.970	0.019
$\sigma_{33} = 0.64$	0.006	0.146	0.955	0.145	0.007	0.140	0.975	0.139
$\sigma_{34} = 0.08$	0.000	0.047	0.935	0.046	0.002	0.045	0.925	0.045
$\sigma_{44} = 0.25$	−0.012	0.032	0.880	0.035	−0.001	0.033	0.920	0.032

reasonably close to 0.95. Under model overparameterization, the estimates of the shape parameters θ_1 and θ_2 from model JM_{BR} are correctly close to zero, suggesting that this model is still a reasonable choice in the absence of outliers.

In simulation setting II, we simulate data from model JM_{BR} with some outliers and extremal events around the tail areas in both longitudinal outcomes by setting the share parameters $\theta_1 = 0.2$ and $\theta_2 = 0.15$. The results suggest that when data come from model JM_{BR}, model JM_{BR} can successfully recover all parameters, including θ_1 and θ_2. In contrast, model JM_{Be} gives more

Table 3. Setting III: simulation results from models JM_{Be} and JM_{BR} when data are simulated from model JM_{Be} with added noise.

	JM_{Be}			JM_{BR}		
	Bias	SD	RMSE	Bias	SD	RMSE
For the first longitudinal outcome						
$\omega_{10} = -1.5$	0.015	0.116	0.116	0.005	0.116	0.116
$\omega_{11} = -0.5$	−0.012	0.151	0.151	−0.005	0.153	0.153
$\beta_{10} = 1.5$	−0.126	0.050	0.136	−0.071	0.053	0.089
$\beta_{11} = -0.5$	0.055	0.065	0.085	0.022	0.069	0.072
$\beta_{12} = -0.1$	0.011	0.012	0.016	0.003	0.013	0.013
$\beta_{13} = 0.2$	−0.019	0.017	0.026	−0.005	0.017	0.018
$\phi_1 = 10$	−2.316	0.264	2.331	0.810	0.377	0.893
For the second longitudinal outcome						
$\omega_{20} = -1$	−0.005	0.086	0.086	−0.004	0.082	0.082
$\omega_{21} = -1$	0.006	0.130	0.130	0.002	0.129	0.129
$\beta_{20} = 1$	−0.101	0.053	0.114	−0.091	0.052	0.105
$\beta_{21} = -1$	0.083	0.071	0.110	0.071	0.068	0.098
$\beta_{22} = -0.2$	0.018	0.015	0.023	0.015	0.015	0.022
$\beta_{23} = 0.5$	−0.036	0.021	0.042	−0.029	0.020	0.035
$\phi_2 = 5$	−0.480	0.133	0.498	0.164	0.287	0.330
For the survival outcome						
$\psi = -1$	−0.043	0.166	0.171	−0.041	0.168	0.172
$\nu_{11} = 0.2$	0.017	0.176	0.177	0.023	0.182	0.183
$\nu_{12} = 0.4$	0.034	0.286	0.287	−0.004	0.246	0.245
$\nu_{21} = 0.5$	0.110	0.253	0.275	0.094	0.246	0.263
$\nu_{22} = 0.4$	0.039	0.425	0.425	0.032	0.414	0.414
For the random effects covariance matrix						
$\sigma_{11} = 1.44$	−0.009	0.266	0.265	0.001	0.265	0.265
$\sigma_{12} = 0.288$	−0.041	0.057	0.070	−0.027	0.060	0.065
$\sigma_{13} = 0.192$	−0.014	0.102	0.103	−0.009	0.101	0.102
$\sigma_{14} = 0.3$	−0.031	0.051	0.059	−0.024	0.052	0.057
$\sigma_{22} = 0.36$	−0.072	0.032	0.079	−0.024	0.032	0.040
$\sigma_{23} = 0.048$	−0.005	0.040	0.041	−0.002	0.041	0.041
$\sigma_{24} = 0.06$	−0.012	0.018	0.022	−0.009	0.019	0.021
$\sigma_{33} = 0.64$	−0.004	0.155	0.154	0.004	0.150	0.149
$\sigma_{34} = 0.08$	−0.007	0.040	0.040	−0.008	0.041	0.042
$\sigma_{44} = 0.25$	−0.043	0.026	0.050	−0.033	0.027	0.042

biased estimates, larger RMSE, and low coverage probabilities, especially for the mean-related regression parameters $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$.

In simulation setting III, we evaluate the influence of outliers and extremal events around the tail areas. This simulation setting is similar to setting I, but we contaminate 1% of the randomly selected observations from both longitudinal outcomes with high scores (0.9, 1) by decreasing Δ units ($y^*_{ijk} = y_{ijk} - \Delta$, and $\Delta = 0.8$). We then fit joint models JM_{Be} and JM_{BR} (Table 3). Comparing with model JM_{Be}, model JM_{BR} provides parameter estimates with smaller bias (e.g. -0.071 versus -0.126 for β_{10}) and smaller RMSE (e.g. 0.089 versus 0.136 for β_0), suggesting that model JM_{BR} can effectively control the outlying observations.

In conclusion, the simulation results suggest that in the absence of outliers and extremal events around the tail areas, models JM_{Be} and JM_{BR} give comparable results after accounting for dependent censoring. In the presence of outliers and extremal events around the tail areas, model JM_{BR} provide more accurate parameter estimates than model JM_{Be}.

6 Application to the LS-1 study

In this section, we apply the proposed joint models (JM_{Be} and JM_{BR}) as well as the reduced models (RM_{Be} and RM_{BR}) to the motivating LS-1 study. For all results in this section, we run two parallel MCMC chains with overdispersed initial values and run each chain for 4000 iterations. The first 2000 iterations are discarded as burn-in and the inference is based on the remaining 2000 iterations from each chain. Good mixing properties of the MCMC chains for all model parameters are observed in the trace plots. The scale reduction $\hat{R}$ of all parameters are smaller than 1.1. To specify the baseline hazard functions $\lambda_0(t_i)$, we divide the time to death or dropout into $M=3$ intervals by every $1/M$th quantiles. We have also explored other selections of M and the results are very similar.

In the data analysis, we divide both the variables EQ-VAS and PDQ-39 by 100 to rescale it to the interval [0, 1]. Because there is no occurrence of 0 in the variable EQ-VAS ($k=1$) and no occurrence of 1 (or 100 in the original scale) in the variable PDQ-39 ($k=2$), we fit the following model:

$$\begin{aligned}
\text{logit}[P(y_{ij1}=1|u_{i1})] &= \mathbf{X}_{i1}\boldsymbol{\omega}_1 + u_{i1}, \quad \text{logit}(\gamma_{ij1}|u_{i2}) = \mathbf{X}_{i2}\boldsymbol{\beta}_1 + u_{i2} \\
\text{logit}[P(y_{ij2}=0|u_{i3})] &= \mathbf{X}_{i1}\boldsymbol{\omega}_2 + u_{i3}, \quad \text{logit}(\gamma_{ij2}|u_{i4}) = \mathbf{X}_{i2}\boldsymbol{\beta}_2 + u_{i4} \\
\lambda(t_i) &= \lambda_0(t_i)\exp[\mathbf{X}_i\boldsymbol{\psi} + \nu_1 u_{i1} + \nu_2 u_{i2} + \nu_3 u_{i3} + \nu_4 u_{i4}]
\end{aligned}$$

where $\mathbf{X}_{i1} = (1, x_i)'$, $\mathbf{X}_{i2} = (1, x_i, t_{ij}, x_i \times t_{ij})'$, and $\boldsymbol{X}_i = x_i$, x_i is the treatment assignment variable (1 for creatine, and 0 for placebo), t_{ij} is the visit time in years, and the random effects vector $\mathbf{u}_i = (u_{i1}, u_{i2}, u_{i3}, u_{i4})' \sim MVN(0, \boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma}$ is a covariance matrix of dimension 4×4 with variance $\sigma_1^2, \ldots, \sigma_4^2$ on the diagonal and correlation coefficients ρ_{mn} with $1 \le m < n \le 4$. Table 4 compares models JM_{Be}, JM_{BR}, RM_{Be} and RM_{BR} using the model selection criteria discussed in Section 4.2. Model JM_{BR} performs significantly better than the other three models with smaller $\overline{D}$, DIC_3, EAIC, EBIC, and larger LPML. All Bayes factors in favor of model JM_{BR} over others are larger than 150, suggesting decisive evidence in favor of model JM_{BR}.[31] Thus, we select model JM_{BR} as the final model. Moreover, both joint models JM_{Be} and JM_{BR} outperform their reduced model counterparts RM_{Be} and RM_{BR}, respectively, which indicates that joint modeling is essential and improves the model fitting.

Table 4. Model comparison statistics for the LS-1 dataset. $\overline{D}$: the posterior mean of the deviance; DIC_3: deviance information criterion; EAIC: expected Akaike information criterion; EBIC: expected Bayesian information criterion; LPML: log-pseudo-marginal likelihood. The best fitting model is highlighted in bold.

	$\overline{D}$	DIC_3	EAIC	EBIC	LPML	BF
JM_{BR}	**−34203.2**	**−37183.7**	**−37115.7**	**−36930.0**	**9253.8**	Reference
JM_{Be}	−34781.2	−31987.9	−34717.2	−34542.4	9195.9	>> 150
RM_{BR}	−36954.8	−33948.0	−36894.8	−36730.9	9193.9	>> 150
RM_{Be}	−34549.6	−31713.8	−34493.6	−34340.7	9124.6	>> 150

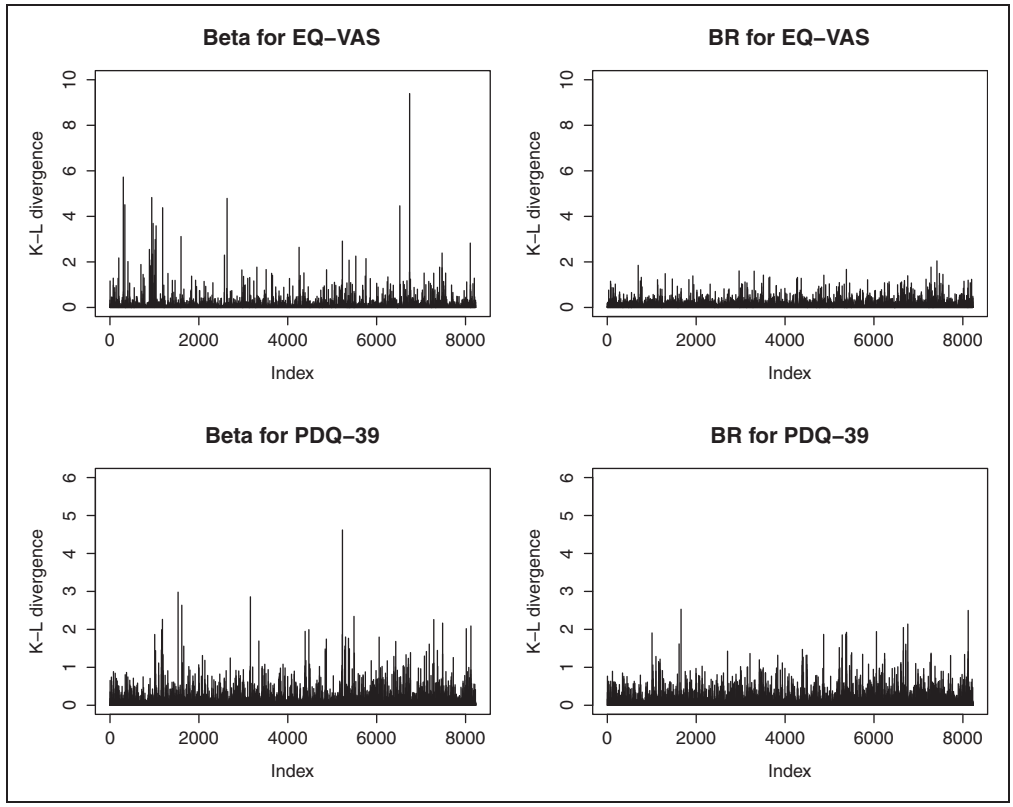

Figure 3. Estimated K-L divergence measures from models JM_{Be} (left panels) and JM_{BR} (right panels) for the longitudinal outcomes EQ-VAS (upper panels) and PDQ-39 (lower panels).

To determine the presence of possible outlying observations, the K-L divergence measure of every observation is presented in Figure 3 for models JM_{Be} (left panels) and JM_{BR} (right panels). Model JM_{Be} identifies many observations as potential outliers, whose K-L divergence measures are larger than 3. However, using model JM_{BR}, there seem to be no influential observations with all K-L divergence measures smaller than 3. This figure suggests that model JM_{BR} can effectively control the potential outlying observations.

Table 5 displays the posterior means, standard deviations (SDs), and the 95% equal-tailed credible intervals from the joint models JM_{Be} and JM_{BR}. Parameter estimates are noticeably different from two models, although the same set of parameters are identified for statistical significance. For the creatine patients, the odds ratio of the probability of reporting 1 (best imaginable health, or 100 in the original scale) in the EQ-VAS score is 0.605 (exp(−0.502), 95% CI: [0.350, 1.040]) comparing with the placebo patients, in model JM_{BR}, versus 0.602 (exp(−0.508), 95% CI: [0.353, 1.034]) in model JM_{Be}. Similarly, the odds ratio of the probability of reporting 0 in the PDQ-39 score (best QoL) for the creatine patients is 0.869 (exp(−0.140), 95% CI: [0.497, 1.534]) comparing with the placebo patients, in model JM_{BR}, versus 0.839 (exp(−0.176), 95% CI: [0.480, 1.554]) in model JM_{Be}.

The parameters under the sections of conditional mean represent the effects of the covariates on the means of the EQ-VAS and PDQ-39 scores conditional on not being on the boundaries. Thus,

Table 5. Results of fitting models JM_{Be} and JM_{BR} in the LS-1 dataset.

	JM_{Be}				JM_{BR}			
	Mean	SD	2.5%	97.5%	Mean	SD	2.5%	97.5%
For EQ-VAS probability of being one								
Int	−6.367	0.370	−7.138	−5.679	−6.336	0.372	−7.135	−5.683
Creatine	−0.508	0.275	−1.042	0.033	−0.502	0.277	−1.051	0.039
For EQ-VAS conditional mean								
Int	1.528	0.026	1.477	1.578	1.550	0.026	1.499	1.601
Creatine	−0.039	0.036	−0.113	0.030	−0.048	0.035	−0.116	0.022
Time	−0.079	0.006	−0.090	−0.067	−0.079	0.005	−0.089	−0.069
Time:Creatine	−0.003	0.008	−0.020	0.013	−0.001	0.007	−0.015	0.013
ϕ_1	14.339	0.252	13.849	14.843	25.131	0.620	23.938	26.342
θ_1					0.110	0.009	0.093	0.128
For PDQ-39 probability of being zero								
Int	−6.991	0.428	−7.908	−6.226	−7.032	0.430	−7.913	−6.223
Creatine	−0.176	0.296	−0.734	0.441	−0.140	0.288	−0.700	0.428
For PDQ-39 conditional mean								
Int	−2.084	0.030	−2.144	−2.026	−2.063	0.030	−2.124	−2.002
Creatine	0.001	0.042	−0.081	0.084	0.005	0.042	−0.079	0.088
Time	0.128	0.005	0.118	0.138	0.126	0.005	0.117	0.136
Time:Creatine	0.019	0.007	0.004	0.033	0.014	0.007	0.001	0.028
ϕ_2	29.786	0.544	28.719	30.879	33.947	0.790	32.384	35.502
θ_2					0.040	0.007	0.027	0.056
For survival outcome								
Creatine	0.030	0.108	−0.183	0.239	0.026	0.106	−0.185	0.232
ν_1	0.055	0.251	−0.433	0.518	0.029	0.193	−0.374	0.401
ν_2	−0.419	1.313	−2.767	2.322	−0.230	0.840	−1.778	1.535
ν_3	0.233	0.498	−0.671	1.438	0.225	0.376	−0.470	1.059
ν_4	1.209	1.779	−1.976	5.404	1.216	1.251	−1.028	3.955
For the random effects covariance matrix								
σ_1	2.693	0.230	2.283	3.156	2.649	0.234	2.246	3.161
σ_2	0.606	0.014	0.580	0.633	0.629	0.013	0.604	0.656
σ_3	2.624	0.229	2.217	3.121	2.621	0.235	2.168	3.095
σ_4	0.783	0.015	0.755	0.813	0.787	0.015	0.758	0.817
ρ_{12}	0.768	0.054	0.659	0.864	0.742	0.051	0.634	0.836
ρ_{13}	0.576	0.072	0.424	0.706	0.586	0.069	0.442	0.712
ρ_{14}	−0.529	0.048	−0.621	−0.436	−0.528	0.052	−0.629	−0.424
ρ_{23}	0.569	0.058	0.456	0.681	0.539	0.056	0.420	0.638
ρ_{24}	−0.727	0.016	−0.757	−0.695	−0.705	0.015	−0.735	−0.675
ρ_{34}	−0.889	0.047	−0.966	−0.786	−0.873	0.048	−0.954	−0.775

negative parameters suggest deterioration in patients' global assessment of their health represented in the EQ-VAS score, while positive parameters suggests deterioration in patients' QoL represented in the PDQ-39 score. Conditional on other covariates and the random effects, parameter interpretation can be expressed in terms of the covariate effect on the odds $\frac{\gamma_{ijk}}{1-\gamma_{ijk}}$ (rescale) or $\frac{100\gamma_{ijk}}{100-100\gamma_{ijk}}$ in the original scale for model JM_{BR}. Specifically, for the creatine patients, the ratio

between the expected EQ-VAS score γ_{ij1} and the difference to perfect health $(1-\gamma_{ij1})$ is 0.953 (exp(−0.048), 95% CI: [0.890, 1.022]) times the ratio of the placebo patients. For one year increase in time, the ratio between the expected EQ-VAS score and the difference to perfect health decreases by 7.6% (1 − exp(−0.078), 95% CI: [0.067, 0.085]). The regression parameters for the PDQ-39 scores can be interpreted in a similar way. The fact that the estimates of the shape parameters θ_1 and θ_2 are 0.110 (95% CI: [0.093, 0.128]) and 0.040 (95% CI: [0.027, 0.056]) suggests the existence of some potential outliers in the EQ-VAS score, but possibly not in the PDQ-39 score. Moreover, creatine has insignificant effects on the risk of death or dropout, that is, the time to death or dropout for the creatine patients is estimated to be 1.026 times (exp(0.026), 95% CI: [0.831, 1.261]) that of the placebo patients. The findings of insignificant creatine effects are consistent with the primary paper of the LS-1 study.[39] There is no significant association between the longitudinal outcomes EQ-VAS and PDQ-39 and the survival time, as indicated by the insignificant parameters ν_1, ν_2, ν_3, and ν_4. The regression parameters in model JM_{Be} can be interpreted in a similar way.

Table 5 also suggests strong correlation both within and between the two longitudinal outcomes EQ-VAS and PDQ-39 as indicated by the relatively large correlation coefficients ρ. Specifically, patients with higher probability of being 1 in the EQ-VAS score (perfect health) are more likely to report high EQ-VAS values (good health) as indicated by large positive significant $\hat{\rho}_{12}$, more likely to report 0 in the PDQ-39 score (best QoL) as indicated by large positive significant $\hat{\rho}_{13}$, and less likely to report high PDQ-39 values (worse QoL) as indicated by large negative significant $\hat{\rho}_{14}$. Similarly, patients with higher probability of reporting high EQ-VAS values are more likely to report 0 in the PDQ-39 score (large positive significant $\hat{\rho}_{23}$) and less likely to report high PDQ-39 values (large negative significant $\hat{\rho}_{24}$). Lastly, patients with higher probability of being 0 in the PDQ-39 score are less likely to report high PDQ-39 values (large negative significant $\hat{\rho}_{34}$).

7 Discussion

Health-related QoL variables in the form of PROs have become increasingly common in clinical trials of Parkinson's disease (PD) and other neurodegenerative diseases (e.g. Alzheimer's disease, Huntington's disease) in the past three decades,[40] e.g. EQ-VAS and PDQ-39 scores in the motivating LS-1 study. When the survival benefit of a treatment is small or modest, improvement in QoL for patients is sometimes more important than modest survival benefit in making treatment decisions.[41] Because there is neither a cure nor a treatment that can effectively slow the progression of PD, it is essential to model longitudinal PROs related to PD patients' QoL and terminal events through a joint modeling framework to gain insight into their correlation. Moreover, some longitudinal PROs such as variables EQ-VAS and PDQ-39 are constrained in the closed interval [0, 1] after rescaling. Although traditional Gaussian models works well for the normally distributed continuous outcomes with no boundary, models based on Beta distributions are more appropriate in this scenario to avoid out of bound predictions. To better analyze the data with multiple PROs in the close unit interval [0, 1], we develop a joint modeling framework to address the following data complexities: (1) multiple correlated PROs; (2) the presence of the boundary values of zeros and ones; (3) extreme outliers and heavy tails; (4) the PROs-dependent terminal events such as death and dropout.

Our modeling framework consists of a multivariate augmented BR regression sub-model for the multiple longitudinal outcomes and a Cox proportional hazard sub-model for the dependent censoring event. Two sub-models are linked by shared random effects denoting subject-specific characteristics. We adopt Bayesian inference framework based on MCMC simulation for parameter estimation. The extensive simulation study suggests that in the presence of outliers and heavy tails, the proposed joint modeling framework based on the BR distributions (model JM_{BR})

improves the accuracy of parameter estimates, while the joint modeling framework based on the Beta distribution (model JM_{Be}) provides parameter estimates with large bias and RMSE and poor coverage probabilities. We apply the proposed models to the motivating LS-1 study dataset. Model JM_{BR} has a better fit than model JM_{Be} (which does not account for the heavy tails and outliers), and the reduced models RM_{Be} and RM_{BR} (which assume independence between the survival time and longitudinal outcomes). The treatment creatine has insignificant effects in the probabilities of being at the boundaries, in the mean scores of both EQ-VAS (global health assessment) and PDQ-39 (quality of health) variables, and in the survival time. The findings of insignificant creatine effects are not surprising because the LS-1 study was terminated early for futility based on results of a planned interim analysis.[39] However, our final model JM_{BR} detects a significant time effect on both EQ-VAS and PDQ-39 variables and a significant time and treatment interaction effect on outcome PDQ-39, which is a new finding and is worth further investigation. Moreover, our model provides unique clinical insight into the correlation between PROs (inter-outcome correlation). Although our data example only contains one boundary value for each outcome, the proposed JM_{BR} model is general and flexible to make it widely applicable to longitudinal outcomes with the support of $[0, 1]$ (either with or without rescaling), in the presence of outcome-dependent survival events, because it can accommodate various data features, e.g. either with or without boundary observations and outliers and heavy tails. The proposed model and Bayesian inference can be easily implemented by the publicly available software packages such as `BUGS` and `STAN`. The generic structure of `BUGS` and `STAN` facilitates flexibility in model specification. We expect that the `STAN` code we provide would help make our model easily accessible to applied researchers.

There are some limitations in our proposed model that we will address in our future study. This article only considers a composite endpoint (death or dropout) as a single type terminal event. The covariate effects for the risks of these two events can be different. The proposed joint model can be extended to accommodate competing risks survival data. Moreover, relaxing the proportionality assumption in the Cox model by exploring other survival model choices such as accelerated failure time model warrants further investigation. Some researchers[42,43] have reported that the statistical inference of joint models is generally robust to the departure from the normality assumption. It is of interest to investigate our joint models' performance when the underlying random effects distribution is symmetric non-normal or even asymmetric. In addition, the random effects variance is assumed to be homogeneous (same for all individuals). However, the random effects variance may depend on subject-specific characteristics and is thus heterogeneous. Ignoring the heterogeneity can result in biased estimates.[44,45] As a future direction, we would address the issues of non-normal, heterogeneous, or even non-parametric random effects[46] in the proposed joint modeling framework.

Acknowledgements

The authors acknowledge the Texas Advanced Computing Center (TACC) at The University of Texas at Austin for providing high-performing computing resources that have contributed to the research results reported within this article. URL: http://www.tacc.utexas.edu. The author thanks the editor and two anonymous referees for their reading and valuable comments.

Funding

The project described was supported by the National Center for Research Resources and the National Center for Advancing Translational Sciences, National Institutes of Health (grant number KL2 TR000370). The

content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

References

1. FDA et al. Guidance for industry: Patient-reported outcome measures: Use in medical product development to support labeling claims. *Federal Register* 2009; 74(235): 65132–65133.
2. Peto V, Jenkinson C, Fitzpatrick R, et al. The development and validation of a short measure of functioning and well being for individuals with Parkinson's disease. *Qual Life Res* 1995; **4**(3): 241–248.
3. Kieschnick R and McCullough B. Regression analysis of variates observed on (0, 1): percentages, proportions and fractions. *Statist Modell* 2003; **3**(3): 193–213.
4. Ferrari S and Cribari-Neto F. Beta regression for modelling rates and proportions. *J Appl Statist* 2004; **31**(7): 799–815.
5. Verkuilen J and Smithson M. Mixed and mixture regression models for continuous bounded responses using the Beta distribution. *J Educ Behav Statist* 2012; **37**(1): 82–113.
6. Figueroa-Zúniga J, Arellano-Valle R and Ferrari S. Mixed beta regression: A Bayesian perspective. *Computat Statist Data Anal* 2013; **61**: 137–147.
7. Galvis D, Bandyopadhyay D and Lachos V. Augmented mixed beta regression models for periodontal proportion data. *Statist Med* 2014; **33**(21): 3759–3771.
8. Hatfield L, Boye M and Carlin B. Joint modeling of multiple longitudinal patient-reported outcomes and survival. *J Biopharm Statist* 2011; **21**(5): 971–991.
9. Hatfield LA, Boye ME, Hackshaw MD, et al. Multilevel Bayesian models for survival times and longitudinal patient-reported outcomes with many zeros. *J Am Statist Assoc* 2012; **107**(499): 875–885.
10. Hahn E. Mixture densities for project management activity times: A robust approach to PERT. *Eur J Operat Res* 2008; **188**(2): 450–459.
11. García C, Pérez JG and van Dorp J. Modeling heavy-tailed, skewed and peaked uncertainty phenomena with bounded support. *Statist Meth Applic* 2011; **20**(4): 463–486.
12. Bayes CL, Bazán JL, García C, et al. A new robust regression model for proportions. *Bayesian Anal* 2012; **7**(4): 841–866.
13. Henderson R, Diggle P and Dobson A. Joint modelling of longitudinal measurements and event time data. *Biostatistics* 2000; **1**(4): 465–480.
14. Tsiatis AA and Davidian M. Joint modeling of longitudinal and time-to-event data: An overview. *Statistica Sinica* 2004; **14**(3): 809–834.
15. Yu M, Law NJ, Taylor JM, et al. Joint longitudinal-survival-cure models and their application to prostate cancer. *Statistica Sinica* 2004; **14**(3): 835–862.
16. Luo S. Joint analysis of stochastic processes with application to smoking patterns and insomnia. *Statist Med* 2013; **32**(29): 5133–5144.
17. He B and Luo S. Joint modeling of multivariate longitudinal measurements and survival data with applications to Parkinsons disease. *Statist Meth Med Res* 2013; in press).
18. Luo S, Su X, DeSantis S, et al. Joint model for a diagnostic test without a gold standard in the presence of a dependent terminal event. *Statist Med* 2014; **33**(15): 2554–2566.
19. Luo S. A Bayesian approach to joint analysis of multivariate longitudinal data and parametric accelerated failure time. *Statist Med* 2014; **33**: 580–594.
20. Elm J, et al. Design innovations and baseline findings in a long-term Parkinson's trial: The National Institute of Neurological Disorders and Stroke Exploratory Trials in Parkinson's Disease Long-Term Study–1. *Movement Disorders* 2012; **27**(12): 1513–1521.
21. Lawless J and Zhan M. Analysis of interval-grouped recurrent-event data using piecewise constant rate functions. *Can J Statist* 1998; **26**(4): 549–565.
22. Feng S, Wolfe R and Port F. Frailty survival model analysis of the national deceased donor kidney transplant dataset using poisson variance structures. *J Am Statist Assoc* 2005; **100**(471): 728–735.
23. Liu L and Huang X. Joint analysis of correlated repeated measures and recurrent events processes in the presence of death, with application to a study on acquired immune deficiency syndrome. *J R Statist Soc Ser C (Appl Statist)* 2009; **58**(1): 65–81.
24. Duane S, Kennedy A, Pendleton B, et al. Hybrid Monte Carlo. *Phys Lett B* 1987; **195**(2): 216–222.
25. Neal R. An improved acceptance procedure for the hybrid Monte Carlo algorithm. *J Computat Phys* 1994; **111**(1): 194–203.
26. Hoffman M and Gelman A. The no-U-turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo. *J Machine Learning Res* 2013; in press).
27. Stan Development Team. *Stan Modeling Language Users Guide and Reference Manual, Version 2.5.0*, 2014.
28. Gelman A, Carlin J, Stern H, et al. *Bayesian Data Analysis*. Boca Raton, FL: CRC Press, 2013.
29. Celeux G, Forbes F, Robert C, et al. Deviance information criteria for missing data models. *Bayesian Anal* 2006; **1**(4): 651–673.
30. Carlin B and Louis T. *Bayesian Methods for Data Analysis*. London: Chapman & Hall/CRC, 2009.
31. Kass R and Raftery A. Bayes factors. *J Am Statist Assoc* 1995; **90**: 773–795.
32. Lewis S and Raftery A. Estimating bayes factors via posterior simulation with the Laplace–Metropolis estimator. *J Am Statist Assoc* 1997; **92**: 648–655.
33. Geisser S. *Predictive Inference*. Vol. 55, Boca Raton, FL: CRC Press, 1993.
34. Dey D, Chen M and Chang H. Bayesian approach for nonlinear random effects models. *Biometrics* 1997; **53**(4): 1239–1252.
35. Sinha D and Dey D. Semiparametric Bayesian analysis of survival data. *J Am Statist Assoc* 1997; **92**(439): 1195–1212.
36. Ghosh P and Hanson T. A semiparametric Bayesian approach to multivariate longitudinal data. *Austral N Z J Statist* 2010; **52**(3): 275–288.
37. Peng F and Dey D. Bayesian analysis of outlier problems using divergence measures. *Can J Statist* 1995; **23**(2): 199–213.
38. Cancho VG, Dey DK, Lachos VH, et al. Bayesian nonlinear regression models with scale mixtures of skew-normal distributions: estimation and case influence diagnostics. *Computat Statist Data Anal* 2011; **55**(1): 588–602.
39. Kieburtz K, Tilley BC, Elm JJ, et al. Effect of creatine monohydrate on clinical progression in patients with Parkinson disease: A randomized clinical trial. *JAMA* 2015; **313**(6): 584–593.

40. Martinez-Martin P and Kurtis MM. Health-related quality of life as an outcome variable in Parkinson's disease. *Ther Adv Neurol Disord* 2012; **5**(2): 105–117.
41. Ibrahim JG, Chu H and Chen LM. Basic concepts and methods for joint models of longitudinal and survival data. *J Clin Oncol* 2010; **28**(16): 2796–2801.
42. Song X, Davidian M and Tsiatis AA. A semiparametric likelihood approach to joint modeling of longitudinal and time-to-event data. *Biometrics* 2002; **58**(4): 742–753.
43. Zeng D and Cai J. Asymptotic results for maximum likelihood estimators in joint analysis of repeated measurements and survival time. *Ann Statist* 2005; **33**(5): 2132–2163.
44. Heagerty PJ and Kurland BF. Misspecified maximum likelihood estimates and generalised linear mixed models. *Biometrika* 2001; **88**(4): 973.
45. Daniels MJ and Zhao YD. Modelling the random effects covariance matrix in longitudinal data. *Statist Med* 2003; **22**(10): 1631–1647.
46. Escobar M. Estimating normal means with a Dirichlet process prior. *J Am Statist Assoc* 1994; **89**(425): 268–277.