

---

# Bayesian quantile regression joint models: inference and dynamic predictions

Journal Title

XX(X):2–20

©The Author(s) 2016

Reprints and permission:

sagepub.co.uk/journalsPermissions.nav

DOI: 10.1177/ToBeAssigned

www.sagepub.com/



Ming Yang<sup>1</sup>, Sheng Luo<sup>2</sup> and Stacia DeSantis<sup>1</sup>

## Abstract

In the traditional joint models (JM) of a longitudinal and time-to-event outcome, a linear mixed model (LMM) assuming normal random errors is used to model the longitudinal process. However, in many circumstances, the normality assumption is violated and the LMM is not an appropriate sub-model in the JM. In addition, as the LMM models the conditional mean of the longitudinal outcome, it is not appropriate if clinical interest lies in making inference or prediction on median, lower, or upper ends of the longitudinal process. To this end, quantile regression (QR) provides a flexible, distribution-free way to study covariate effects at different quantiles of the longitudinal outcome and it is robust not only to deviation from normality, but also to outlying observations. In this article, we present and advocate the linear quantile mixed model (LQMM) for the longitudinal process in the JM framework. Our development is motivated by a large prospective study of Huntington's Disease (HD) where primary clinical interest is in utilizing longitudinal motor scores and other early covariates to predict the risk of developing HD. We develop a Bayesian method based on the location-scale representation of the asymmetric Laplace distribution (ALD), assess its performance through an extensive simulation study, and demonstrate how this LQMM-based JM approach can be used for making subject-specific dynamic predictions of survival probability.

## Keywords

Asymmetric Laplace distribution, Bayesian inference, Huntington's disease, Linear quantile mixed model, MCMC

## 1 Introduction

Joint models (JM) of longitudinal and time-to-event data have been well developed in the literature as they are highly applicable in the settings of longitudinal data analysis with event-related terminal events (e.g., dropout or death which is related to the longitudinal process), or of survival analysis with time-dependent covariates measured with error.<sup>1,2</sup> JM have been extensively studied in recent years and an excellent review is provided by Tsiatis and Davidian.<sup>3</sup>

Within the traditional JM framework, a linear mixed model (LMM) is used to model the longitudinal continuous outcome. To account for within-subject correlation, measurements from the same individual share random effects, and random errors are assumed to be normally distributed.<sup>4</sup> However, if normality assumption of the random errors is violated (even after applying various outcome transformations), an LMM is not appropriate. Further, an LMM models covariate effects on the conditional mean of the longitudinal outcome. However, in many clinical studies it is of interest to make inference or prediction on median, lower or higher quantiles of the longitudinal outcome. For example, in a study of the impact of various health factors on low birth weight, which is closely associated with infant mortality and development of chronic diseases, researchers fit a quantile regression (QR) model and find several significantly different covariate effects on lower birth weight compared with that on the mean birth weight.<sup>5</sup> Thus, QR may serve as an alternative to conditional mean regression when the aforementioned limitations are present, or when the subject matter dictates.

In contrast to mean regression such as LMMs, original QR models offer a flexible framework that relaxes the distributional assumption, and provides a way to study covariate effects on various quantiles of the outcome.<sup>6</sup> QR has attracted much attention since the seminal work of Koenker and Bassett Jr.,<sup>7</sup> and extensions for longitudinal data have been explored by many authors.<sup>8–11</sup> Recently, Farcomeni and Viviani<sup>12</sup> first incorporated a linear quantile mixed model (LQMM) into a JM for longitudinal and survival processes, for which they utilized a Monte Carlo Expectation Maximization (MCEM) algorithm for parameter estimation.

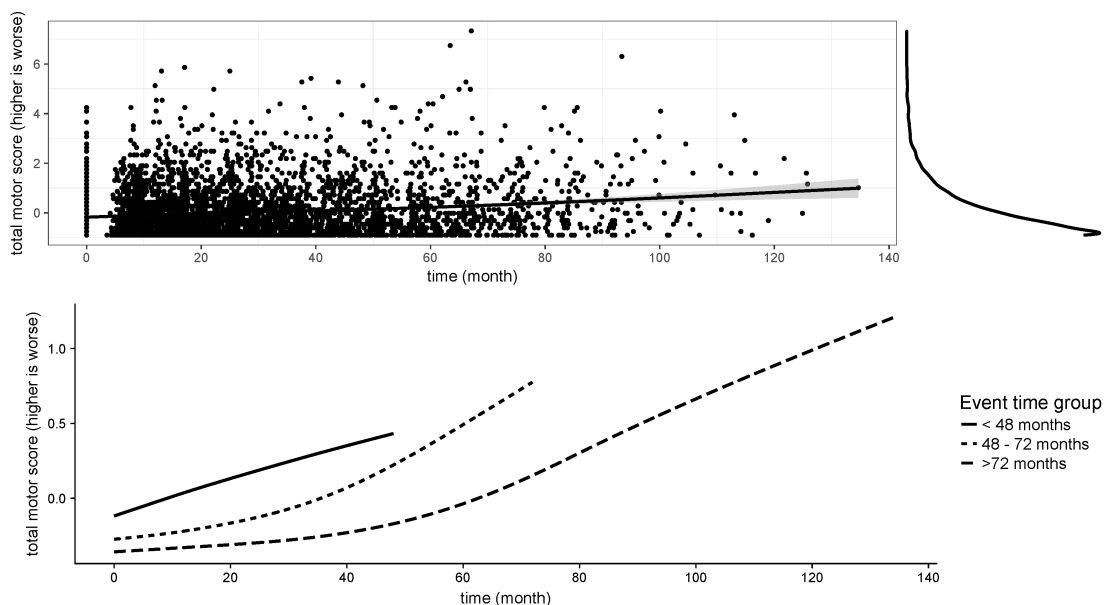
---

<sup>1</sup>Department of Biostatistics, The University of Texas Health Science Center at Houston, 1200 Pressler St, Houston, TX 77030, USA <sup>2</sup> Department of Biostatistics and Bioinformatics, Duke University Medical Center, 2400 Pratt St, 7040 North Pavilion, Durham, NC 27705, USA

**Corresponding author:**

Sheng Luo, Department of Biostatistics and Bioinformatics Duke University Medical Center, 2400 Pratt St, 7040 North Pavilion, Durham, NC 27705, USA.  
Email: sheng.luo@duke.edu

Our work is motivated by the Neurobiological Predictors of Huntington's Disease (HD) study (PREDICT-HD; ClinicalTrials.gov number NCT00051324), a prospective observational study (n=1078) designed to detect early neurobiological predictors of HD. A total of 40 longitudinal biomarkers were measured during the study follow-up of 12 years, where the outcome of interest was time to motor diagnosis of HD (referred to as HD onset). As the primary focus of this study was to measure the association between longitudinal biomarkers and risk of HD onset, the clinical question lends itself to a joint modeling approach. The top panel of Figure 1 displays the scatter plot and kernel density (on the right side) of total motor score (TMS), a commonly used rating criteria of body motion abilities based on the Unified Huntington Disease Rating Scale (UHDRS). The top panel suggests that TMS increases (deteriorates) as time progresses. The bottom panel of Figure 1 displays the mean TMS values over time for individuals with follow-up time less than 48 months (solid line), 48–72 months (dotted line), and more than 72 months (slashed line). It indicates that individuals with shorter follow-up tend to have higher TMS values (worse mobility). Because individuals with higher TMS are at greater risk of developing HD, it is of particular interest to explore the disease progression in a higher TMS subpopulation as opposed to modeling the conditional mean of TMS.



**Figure 1.** Top panel: Scatter plot (with loess curve) and kernel density plot (right side) for total motor score from the study population (time unit: month; lower total motor score is better); bottom panel: Mean total motor score values over time by event time group.

Moreover, we are specifically interested in estimating the risk of developing HD for those individuals who are still free of the disease. To this end, the JM framework offers a novel way of making such personalized dynamic predictions of future event-free probability.<sup>13,14</sup> A key feature of these dynamic predictions frameworks is that the predictive measures can be dynamically updated as additional longitudinal measurements from the target individual become available, providing instantaneous risk assessment. To make dynamic predictions in the current context, we first build the quantile regression joint model (QRJM) consisting of a LQMM for the longitudinal process and a Cox proportional hazard model (PHM) for the time to HD onset process and then draw Bayesian inference, based on a large study population. Next, conditioning on an individual's longitudinal biomarker trajectory information, we make prediction of his/her future HD-free probability by sampling from the posterior distributions of the fixed effects and of the subject-specific random effects. When new longitudinal measurements are available, we dynamically update the ~~his/her~~ HD-free probability. Our work is different from Farcomeni and Viviani<sup>12</sup> in the following: (1) we consider a fully Bayesian QRJM for statistical inference while they used a MCEM algorithm; (2) more importantly, by taking advantage of the posterior distributions of model parameters and subject-specific random effects, we develop a dynamic predictions procedure for future event-free probability under the proposed QRJM.

The rest of this article proceeds as follows. In Section 2, we give details of the QRJM and statistical methods used for inference and dynamic predictions. In Section 3, we present two simulation studies to validate the proposed methods. In Section 4, we apply the proposed methods to the motivating PREDICT-HD study. We conclude the article with a discussion in Section 5.

## 2 Methods

### 2.1 Bayesian Linear Quantile Mixed Model

Let  $Y_i(t_{ij})$  be the longitudinal outcome for individual  $i$  measured at time  $t_{ij}$  where  $i = 1, \dots, N$  and  $j = 1, \dots, n_i$ . Consider the linear mixed effects model:

$$Y_i(t) = \mathbf{X}_i^\top(t)\boldsymbol{\beta} + \mathbf{Z}_i^\top(t)\mathbf{u}_i + \varepsilon_i(t), \varepsilon_i(t) \sim N(0, \sigma^2), \quad (1)$$

where  $\mathbf{X}_i(t)$  is a  $p$ -dimensional covariates,  $\boldsymbol{\beta}$  is the corresponding fixed effect vector,  $\mathbf{Z}_i(t)$  is a  $k$ -dimensional covariates, and  $\mathbf{u}_i$  is the corresponding multivariate normal random effects vector.

A linear quantile mixed model (LQMM) assumes that the conditional quantile of the outcome is a linear function of the covariates,

$$Q_{Y_i(t)|\mathbf{X}_i(t), \mathbf{Z}_i(t)}(\tau) = \mathbf{X}_i^\top(t)\boldsymbol{\beta}_\tau + \mathbf{Z}_i^\top(t)\mathbf{u}_i, \quad (2)$$

where the  $\tau$ th quantile of a random variable  $Y$  is defined as  $Q_Y(\tau) = F_Y^{-1}(\tau) = \inf \{y : F_Y(y) \geq \tau\}$  for  $\tau \in [0, 1]$ . The quantile regression estimates can be obtained by minimizing the following loss function,  $\hat{\beta}_\tau = \arg \min_{\beta \in \mathbb{R}^p} \sum_{i,t} \left[ \rho_\tau \left( Y_i(t) - \mathbf{X}_i^\top(t) \beta_\tau - \mathbf{Z}_i^\top(t) \mathbf{u}_i \right) \right]$ , where  $\rho_\tau(\cdot)$  is defined as  $\rho_\tau(Y) = Y(\tau - I(Y < 0))$ . In quantile regression, parameter estimators are functions of the quantile. So parameter  $\beta_\tau$  is a function of quantile  $\tau$ , as denoted by the subscript.

As discussed in Koenker and Machado<sup>15</sup> and Yu and Moyeed<sup>16</sup>, the above minimization problem can be rephrased as a maximum-likelihood problem by assuming that the random error  $\varepsilon_i(t)$  in (1) follows the asymmetric Laplace distribution (ALD), denoted by  $ALD(0, \sigma, \tau)$  with location parameter equals 0, scale parameter  $\sigma > 0$  and skewness parameter  $\tau \in (0, 1)$ .  $ALD(0, \sigma, \tau)$  is skewed to left when  $\tau > 0.5$ , and skewed to right when  $\tau < 0.5$ . When  $\tau = 0.5$ , ALD reduces to the symmetric Laplace distribution. To visualize this, Web Figure 1 displays the density functions of a standard normal distribution, a symmetric Laplace distribution, and two ALDs with  $\tau = 0.75$  and  $\tau = 0.25$ , respectively. Adopting the ALD, the LQMM in (2) becomes  $Y_i(t) = \mathbf{X}_i^\top(t) \beta_\tau + \mathbf{Z}_i^\top(t) \mathbf{u}_i + \varepsilon_i(t)$ ,  $\varepsilon_i(t) \sim ALD(0, \sigma, \tau)$ , where  $i = 1, \dots, N$  and  $t = 1, \dots, n_i$ . The conditional likelihood function is  $\ell(Y_i(t)|\beta_\tau, \mathbf{u}_i, \sigma) = \frac{\tau(1-\tau)}{\sigma} \exp \left[ -\rho_\tau \left( \frac{Y_i(t) - \mathbf{X}_i^\top(t) \beta_\tau - \mathbf{Z}_i^\top(t) \mathbf{u}_i}{\sigma} \right) \right]$ .

Linear programming algorithms can be applied to obtain parameter estimates under the frequentist framework. However, to develop a Bayesian sampling algorithm for model inference, we utilize a location-scale mixture representation of the ALD,<sup>17</sup> which is a functional form with a mixture of common distributions. Under this parameterization, the random error is represented as  $\varepsilon_i(t) = \kappa_1 e_i(t) + \kappa_2 \sqrt{\sigma e_i(t)} v_i(t)$  with  $v_i(t) \sim \mathcal{N}(0, 1)$ ,  $e_i(t) \sim \exp(1/\sigma)$ ,  $\kappa_1 = \frac{1-2\tau}{\tau(1-\tau)}$ , and  $\kappa_2^2 = \frac{2}{\tau(1-\tau)}$ .

This reparameterization leads to the following LQMM,

$$Y_i(t) = \mathbf{X}_i^\top(t) \beta_\tau + \mathbf{Z}_i^\top(t) \mathbf{u}_i + \kappa_1 e_i(t) + \kappa_2 \sqrt{\sigma e_i(t)} v_i(t), \quad (3)$$

or equivalently, the conditional likelihood function is

$$\ell(Y_i(t)|\beta_\tau, \mathbf{u}_i, e_i(t), \sigma) = \frac{1}{\sqrt{2\pi\kappa_2^2\sigma e_i(t)}} \exp \left[ -\frac{(Y_i(t) - \mathbf{X}_i^\top(t) \beta_\tau - \mathbf{Z}_i^\top(t) \mathbf{u}_i - \kappa_1 e_i(t))^2}{2\kappa_2^2\sigma e_i(t)} \right]. \quad (4)$$

As discussed in Yu and Moyeed<sup>16</sup>, irrespective of the actual distribution of the data, Bayesian quantile regression using ALD distribution works quite well for different error distributions and the performance is quite robust and satisfactory.

## 2.2 Joint Models Using Longitudinal Quantile Regression

We then extend the regular joint models (consisting of a linear mixed sub-model for the longitudinal process and a Cox PHM sub-model for the survival process, referred to as LMJM), by replacing the linear mixed sub-model with an LQMM as in (3). Let  $T_i = \min(T_i^*, C_i)$  be the observed event time for individual  $i$ , where  $T_i^*$  is the true underlying event time and  $C_i$  is the censoring time. Let  $\Delta_i$  be the event indicator (1 if the event is observed, and 0 otherwise). Let  $Y_i(t)$  be the continuous longitudinal outcome for individual  $i$  measured at time  $t$ . Note that  $Y_i(t)$  is only observed when  $t \leq T_i$ , and the complete longitudinal measurements for individual  $i$  can be written as  $\mathcal{Y}_i(t) = \{Y_i(s) : 0 \leq s \leq t\}$ . We denote the true underlying longitudinal measurement for individual  $i$  at time  $t$  with  $m_i(t)$  and his/her complete history of true longitudinal process as  $\mathcal{M}_i(t) = \{m_i(s) : 0 \leq s \leq t\}$ . The proposed quantile regression joint models (QRJM) can be written as a set of two sub-models:

$$\begin{cases} Y_i(t) = m_i(t) + \varepsilon_i(t) = \mathbf{X}_i^\top(t)\boldsymbol{\beta}_\tau + \mathbf{Z}_i^\top(t)\mathbf{u}_i + \varepsilon_i(t), \varepsilon_i(t) \sim ALD(0, \sigma, \tau) \\ h(T_i|\mathcal{M}_i(T_i), \mathbf{W}_i; \gamma_\tau, \alpha_\tau) = h_0(T_i) \exp(\mathbf{W}_i^\top \gamma_\tau + \alpha_\tau (\mathbf{X}_i^\top(T_i)\boldsymbol{\beta}_\tau + \mathbf{Z}_i^\top(T_i)\mathbf{u}_i)), \end{cases} \quad (5)$$

where the first sub-model is the LQMM introduced in Section 2.1, in which  $\mathbf{X}_i(t)$  are the fixed effect covariates and  $\mathbf{Z}_i(t)$  are the covariates associated with  $k$ -dimensional random effects  $\mathbf{u}_i$ . The second sub-model takes the format of PHM where  $h_0(\cdot)$  is the baseline hazard function and  $\mathbf{W}_i$  are the time-independent  $q$ -dimensional fixed effect covariates. These two sub-models are linked by incorporating  $m_i(t)$  (the true underlying longitudinal measurement at time  $t$ ) in the time-to-event process. The association parameter  $\alpha_\tau$  quantifies the strength of association between  $m_i(t)$  and the hazard for event at the same time point, e.g., positive  $\alpha_\tau$  indicates that individuals with higher measurements tend to have an event earlier. Further extension of the JM in the functional form of the two processes is also possible, as discussed in Rizopoulos *et al.*<sup>18</sup>

In the proposed QRJM (5), all parameters are functions of quantile  $\tau$ . Thus, by choosing different quantiles, one can conduct a comprehensive analysis of the relationship between the outcome and the covariates. Depending on the research aims, we can take different strategies to utilize the flexibility of the QRJM. For example, to conduct a study over the entire conditional distribution of the longitudinal outcome, we can fit the QRJM through a set of selected quantiles, collect and compare the resulting parameter estimations. Less varying values in the parameter estimates indicates a relatively stable covariate effects on the outcomes, and vice versa. On the other hand, the interest may lie only in assessing the effect on some pre-specified quantiles (median, lower or higher quantile) of the longitudinal outcome and its association with the time-to-event process.

### 2.3 The Survival Sub-model

For individual  $i$ , the likelihood function for the survival data is:

$$\ell(T_i, \Delta_i | \mathbf{u}_i) = h(T_i | \mathcal{M}_i(T_i), \mathbf{W}_i)^{\Delta_i} S(T_i | \mathcal{M}_i(T_i), \mathbf{W}_i), \quad (6)$$

where  $h(T_i | \mathcal{M}_i(T_i), \mathbf{W}_i)$  is given in (5) and  $S(\cdot)$  is the survival function, and

$$S(T_i | \mathcal{M}_i(T_i), \mathbf{W}_i) = \exp \left\{ - \int_0^{T_i} h_0(s) \exp(\mathbf{W}_i^\top \boldsymbol{\gamma}_\tau + \alpha_\tau (\mathbf{X}_i^\top(s) \boldsymbol{\beta}_\tau + \mathbf{Z}_i^\top(s) \mathbf{u}_i)) ds \right\}.$$

For the baseline hazard  $h_0(t)$ , a parametric form such as exponential model can be used or it can be left unspecified. Specifically, we consider the piecewise-constant baseline hazard function, based on which a closed form survival function can be derived for each time interval. Although all parameters in the proposed QRJM are quantile dependent, for notational simplicity and without ambiguity, we omit the subscript  $\tau$  from all parameters in the following sections (e.g.,  $\boldsymbol{\theta}$  stands for  $\boldsymbol{\theta}_\tau$  for all quantile-based parameters).

### 2.4 Complete Likelihood Function and Bayesian Inference

For individual  $i$ , the complete joint likelihood of the longitudinal and survival data can be written as

$$L_i(\boldsymbol{\theta}; T_i, \Delta_i, \mathcal{Y}_i(T_i), \mathbf{u}_i) = \ell(\mathcal{Y}_i(T_i) | \mathbf{u}_i) \ell(T_i, \Delta_i | \mathbf{u}_i) f(\mathbf{u}_i | \boldsymbol{\Sigma}), \quad (7)$$

where vector  $\boldsymbol{\theta}$  represents all the parameters in (7),  $\ell(\mathcal{Y}_i(T_i) | \mathbf{u}_i) = \prod_{0 \leq t \leq T_i} \ell(Y_i(t) | \mathbf{u}_i)$ ,  $\ell(Y_i(t) | \mathbf{u}_i)$  is given in (4), and  $\ell(T_i, \Delta_i | \mathbf{u}_i)$  is given in (6).

Parameter estimation can be made using Monte Carlo EM (MCEM) algorithm, in which random effects are treated as missing data<sup>12</sup>. However, we take advantage of the location-scale mixture representation of the ALD described in Section 2.1 and propose a fully Bayesian inference approach for parameter estimation and personalized dynamic predictions. Given the complete likelihood in (7) and by Bayes theorem, the posterior distributions of the model parameters are given by

$$f(\boldsymbol{\theta} | \mathbf{T}, \boldsymbol{\Delta}, \mathcal{Y}, \mathbf{u}) \propto \prod_{i=1}^N L_i(T_i, \Delta_i, \mathcal{Y}_i(T_i), \mathbf{u}_i; \boldsymbol{\theta}) f(\boldsymbol{\theta}), \quad (8)$$

where  $\mathbf{T} = \{T_i\}$ ,  $\mathcal{Y} = \{\mathcal{Y}_N(T_i)\}$ ,  $\boldsymbol{\Delta} = \{\Delta_i\}$ ,  $\mathbf{u} = \{\mathbf{u}_i\}$ , for  $i = 1, \dots, N$  and  $f(\boldsymbol{\theta}) = \pi(\boldsymbol{\beta})\pi(\boldsymbol{\gamma})\pi(\boldsymbol{\alpha})\pi(\boldsymbol{\sigma})\pi(\boldsymbol{\Sigma})$  is the product of the prior distributions, where  $\boldsymbol{\Sigma}$  is a  $k \times k$  covariance matrix of the multivariate normal random effects distribution. We adopt the following prior distributions:  $\boldsymbol{\beta} \sim$

$\mathcal{N}_p(\mathbf{0}, 10^3 \mathbf{I})$ ,  $\gamma \sim \mathcal{N}_q(\mathbf{0}, 10^3 \mathbf{I})$ ,  $\alpha \sim \mathcal{N}(0, 10^3)$ ,  $\sigma \sim \mathcal{IG}(10^{-3}, 10^{-3})$ ,  $\Sigma^{-1} \sim \text{Wishart}(\mathbf{I}, k + 1)$ . We also consider the Cholesky decomposition prior for  $\Sigma$  in our simulation studies and find similar results as Wishart prior gives (results not shown). We have investigated other selections of vague prior distributions with various hyper-parameters and have obtained very similar results.

The advantages of using fully Bayesian approach include that the uncertainty of the parameter estimates is fully captured in the posterior distributions and no asymptotic theory is needed to derive the standard error. The fully Bayesian approach provides a straightforward framework to make subject-specific prediction of survival probability using the posterior samples of the parameters and of the posterior predictive distributions for the random effects. Moreover, the proposed QRJM can be readily implemented in JAGS software<sup>19</sup> and the codes have been posted at the Web Supplement to facilitate easy reading and implementation of the proposed QRJM model.

## 2.5 Predictions of Survival Probability

Upon fitting the QRJM to a training dataset with  $N$  individuals, we can make prediction of survival probability for a new individual based on a set of his or her historical longitudinal measurements (denoted by  $\mathcal{Y}_i(t)$ ) as well as other covariates information. Given that the individual is event-free up to the censoring time  $t$ , the conditional survival probability up to time  $m = t + \Delta t$  ( $\Delta t > 0$ ), where  $\Delta t$  the prediction time interval, is denoted as  $p_i(m|t) = \Pr(T_i^* \geq m | T_i^* > t, \mathcal{Y}_i(t); \theta)$ , which can be further elaborated as follows:

$$\begin{aligned} & \Pr(T_i^* \geq m | T_i^* > t, \mathcal{Y}_i(t); \theta) \\ &= \int \Pr(T_i^* \geq m | T_i^* > t, \mathcal{Y}_i(t), \mathbf{u}_i; \theta) f(\mathbf{u}_i | T_i^* > t, \mathcal{Y}_i(t), \theta) d\mathbf{u}_i \\ &= \int \Pr(T_i^* \geq m | T_i^* > t, \mathbf{u}_i; \theta) f(\mathbf{u}_i | T_i^* > t, \mathcal{Y}_i(t), \theta) d\mathbf{u}_i \\ &= \int \frac{S_i[m | \mathcal{M}_i(m, \mathbf{u}_i, \theta); \theta]}{S_i[t | \mathcal{M}_i(t, \mathbf{u}_i, \theta); \theta]} f(\mathbf{u}_i | T_i^* > t, \mathcal{Y}_i(t), \theta) d\mathbf{u}_i, \end{aligned} \tag{9}$$

where  $S(\cdot)$  is the survival function conditional on the entire longitudinal history  $\mathcal{M}_i(\cdot)$  and  $f(\mathbf{u}_i | T_i^* > t, \mathcal{Y}_i(t), \theta) \propto f(T_i^* > t, \mathcal{Y}_i(t), \mathbf{u}_i | \theta) = f(T_i^* > t | \mathbf{u}_i, \theta) f(\mathcal{Y}_i(t) | \mathbf{u}_i, \theta) f(\mathbf{u}_i | \theta)$  is the posterior distribution of the random effects.

To estimate (9), we can use the proposed Bayesian sampling algorithm in Section 2.4 to calculate the posterior mean of the prediction  $E_{\theta|\mathcal{D}_N}[p_i(m|t)]$  and

$$E_{\theta|\mathcal{D}_N}[p_i(m|t)] = \Pr(T_i^* \geq m | T_i^* > t, \mathcal{Y}_i(t), \mathcal{D}_N) = \int \Pr(T_i^* \geq m | T_i^* > t, \mathcal{Y}_i(t); \theta) p(\theta | \mathcal{D}_N) d\theta,$$



where  $\mathcal{D}_N = \{T_i, \Delta_i, \mathbf{Y}_i, i = 1, \dots, N\}$  denotes the training data of size  $N$  and the first part of the equation is given in (9).

A Monte Carlo (MC) estimate of  $p_i(m|t)$  can be obtained using the following procedure:

1. Draw  $\boldsymbol{\theta}^{(p)} \sim Pr(\boldsymbol{\theta}|\mathcal{D}_N)$  for  $p = 1, \dots, P$ ;
2. For each  $\boldsymbol{\theta}^{(p)}$ , draw  $\mathbf{u}_i^{(q)} \sim f(\mathbf{u}_i|T_i^* > t, \mathcal{Y}_i(t), \boldsymbol{\theta}^{(p)})$  for  $q = 1, \dots, Q$  and compute

$$p_i^{(p)}(m|t) = \frac{1}{Q} \sum_{q=1}^Q S_i[m|\mathcal{M}_i(m, \mathbf{u}_i^{(q)}, \boldsymbol{\theta}^{(p)}); \boldsymbol{\theta}^{(p)}] S_i[t|\mathcal{M}_i(t, \mathbf{u}_i^{(q)}, \boldsymbol{\theta}^{(p)}); \boldsymbol{\theta}^{(p)}]^{-1};$$

3. Approximate  $p_i(m|t)$  by  $\hat{p}_i(m|t) = \frac{1}{P} \sum_{p=1}^P p_i^{(p)}(m|t)$  after collecting all  $P$  samples of  $p_i^{(p)}(m|t)$ .

In above algorithm,  $P$  is the total number of post-burnin MC iterations,  $f(\boldsymbol{\theta}|\mathcal{D}_N)$  is the posterior distributions of  $\boldsymbol{\theta}$  given in (8). When an individual does not experience any event and is followed to a further time  $t'$  ( $t' > t$ ), the posterior samples of  $\mathbf{u}_i$  are then drawn from  $f(\mathbf{u}_i|T_i^* > t', \mathcal{Y}_i(t'), \boldsymbol{\theta}^{(p)})$  to reflect the update in data. By following this procedure, the predictions of event-free probability are dynamically updated with new data. The posterior predictive values of the random effects  $\mathbf{u}_i$  are direct results from the MCMC iterations if the individual is from the training dataset. For a new individual who is not in the training dataset, we can use the inference results to run additional MCMC iterations to obtain samples for the new individual's random effects  $\mathbf{u}_i$  and the rest of the algorithm follows. Because each individual only has a few random effects (two in our current model) to estimate, a short MCMC with 200 iterations should be sufficient for convergence.<sup>14</sup> And the uncertainty of the predictions is captured in the sample variance.

In this paper, predictive accuracy from different models is compared from two perspectives, calibration (how close the model predictions are to the true values) and discrimination (how well the models discriminate between individuals who had the event from those who did not). Calibration is only feasible in simulation studies, where the true survival probabilities can be directly calculated from the simulated data. While in data application, time-dependent receiver operating characteristic (ROC) curve method proposed by Heagerty *et al*<sup>20</sup> is used to assess a model's discriminative ability at different prediction time points by calculating the area under the ROC curve (AUC). Higher value in AUC indicates better discriminative ability.

### 3 Simulation Studies

We conduct two simulation studies to validate the proposed QRJM. In the first simulation study, we assess the performance of the proposed Bayesian method in terms of bias and precision of the parameter

estimates. In the second simulation study, we assess the predicted survival probability by comparing with the “gold standard” calculated based on the true parameters and the simulated values of the random effects. In addition, we compare the true time-dependent AUCs with those generated from model predictions.

### 3.1 Simulation Study I: Inferential Performance

In this simulation study, we consider different simulation scenarios where the random errors are generated from either a standard normal distribution or ALD distributions at different quantile  $\tau$ . The simulated data are then fitted using our proposed QRJM (assuming ALD for the random errors) as well as the LMJM (assuming normality for the random errors).

We let the covariate vectors in Model (5) be  $\mathbf{X}_i(t) = (1, x_{i1}, x_{i2} \cdot t)^\top$ ,  $\mathbf{Z}_i(t) = (1, t)^\top$ , and  $\mathbf{W}_i = (w_{i1}, w_{i2})^\top$  with covariates  $x_{i1}, x_{i2}, w_{i1}$  and  $w_{i2}$  being generated from independent standard normal distributions. We simulate the random effects,  $\mathbf{u}_i$ , from a bivariate normal distribution with mean vector  $\mathbf{0}$ , and both standard deviations equal to 0.3 and correlation coefficient equals to 0.16.

To simulate the survival time, we choose constant baseline hazard. We obtain event time  $T_i$  by inverting the survival function after generating  $n$  random values from the standard uniform distribution. We generate the censoring time  $C_i$  from Beta(4, 1) to obtain a censoring proportion around 25%. The longitudinal data are simulated from either a standard normal distribution or a ALD with the location parameter being  $\beta^\top \mathbf{X}_i(t) + \mathbf{u}_i^\top \mathbf{Z}_i(t)$  and dispersion parameter being  $\sigma = 1$ . We keep a maximum of 6 observations for each individual, at follow-up time  $t = (0, 0.25, 0.5, 0.75, 1, 3)$  respectively, after incorporating the time-to-event information.

We consider the following three scenarios in both simulation studies:

1. Scenario 1: random errors follow the ALD with  $\tau = 0.25$  (right-skewed);
2. Scenario 2: random errors follow the ALD with  $\tau = 0.5$  (symmetric about 0 with heavy tails);
3. Scenario 3: random errors follow a standard normal distribution (symmetric about 0).

In each scenario, we simulate 200 datasets with  $N = 600$  in each. We then randomly select 500 individuals as the training data to build the model and use the remaining 100 individuals as the validation data to make out-of-sample predictions.

We report bias, standard error (SE), mean squared error (MSE), and coverage probability (CP) for the QRJM and LMJM. Table 1 suggests that in Scenario 1, the true model (QRJM with  $\tau = 0.25$ ) provides parameter estimates with very small biases and CP being close to the nominal level. In comparison, the QRJM with  $\tau = 0.5$  provides reasonable estimates for most parameters, except the intercept  $\beta_0$ . The poor estimate of  $\beta_0$  (large bias and CP being far from 0.95) should not be surprising because of the incorrect specification of quantile  $\tau$ . The LMJM results in very biased estimates and the CP being away from

the nominal value 0.95 (especially in regression parameters  $\beta$  in the longitudinal model). In Scenario 2 (see Web Table 1) when data are symmetrical about 0 with heavier tails than the normal distribution, the LMJM still produces notably larger bias and lower CP as compared to the true model QRJM with  $\tau = 0.5$ . In Scenario 3 (see Web Table 2), median regression (the QRJM with  $\tau = 0.5$ ) performs comparably to the true model LMJM, suggesting that the QRJM can provide reasonable estimates even when the random errors follow a standard normal distribution. For completion, we also consider a scenario where the random errors are simulated from ALD with  $\tau = 0.75$  (left-skewed). The results (not presented) are very similar to Scenario 1, i.e., the true model QRJM with  $\tau = 0.75$  provides reasonable estimates while LMJM gives biased estimates and CP being far away from 0.95.

**Table 1.** Simulation results in Simulation I Scenario 1 in which random errors are generated from ALD with  $\tau = 0.25$ .

	QRJM ( $\tau = 0.25$ )				QRJM ( $\tau = 0.5$ )				LMJM			
	Bias	SE	MSE	CP	Bias	SE	MSE	CP	Bias	SE	MSE	CP
Coefficients for longitudinal process												
$\beta_0$	-0.003	0.080	0.014	0.930	1.659	0.129	2.807	0.020	2.702	0.146	7.350	0.000
$\beta_1$	0.015	0.068	0.010	0.950	0.024	0.105	0.043	0.890	0.080	0.116	0.052	0.860
$\beta_2$	0.016	0.083	0.013	0.950	0.014	0.112	0.042	0.970	0.078	0.128	0.052	0.920
Coefficients for survival process												
$\gamma_1$	0.005	0.055	0.006	0.940	0.008	0.057	0.006	0.960	0.009	0.058	0.007	0.960
$\gamma_2$	0.006	0.055	0.006	0.930	0.010	0.056	0.007	0.910	0.010	0.058	0.007	0.940
$\alpha$	-0.004	0.078	0.010	0.970	-0.051	0.119	0.070	0.930	-0.087	0.103	0.040	0.800

### 3.2 Simulation Study II: Predictive Performance

In this simulation study, we make predictions for 100 individuals in the validation dataset (out-of-sample predictions) in the three scenarios in Section 3.1. For each individual, we use the simulated data, random effects and the true parameter values to calculate the true survival probability given by  $\frac{S_i[m|\mathcal{M}_i(m, \mathbf{u}_i, \boldsymbol{\theta}); \boldsymbol{\theta}]}{S_i[t|\mathcal{M}_i(t, \mathbf{u}_i, \boldsymbol{\theta}); \boldsymbol{\theta}]}$  and we use it as the “gold standard”.

To assess the prediction validation (how well the models predict the survival probability), we calculate the MSE of the predictions against the “gold standard”. Time-dependent AUC is also calculated for both gold standard survival probabilities (true AUC) as well as the model predictions (predicted AUC) as a measure of model discrimination ability. To make the predictions “dynamic”, we choose different combinations of censoring time (i.e.,  $t$ ) and the prediction time interval (i.e.,  $\Delta t$ ) to mimic expected real-world time points based on our HD dataset. Table 2 summarizes the MSE as well as the time-dependent AUC for three chosen censoring time points ( $t = 0.25, 0.5, 0.75$ ) in Scenario 1. For all combinations of  $(t, \Delta t)$ , the true QRJM with  $\tau = 0.25$  provides accurate estimates of  $p_i(m|t)$  with negligible MSEs

and the estimated AUCs being close to the true AUCs. In comparison, QRJM with  $\tau = 0.5$  (i.e., median regression model) produces larger MSEs but comparable discrimination (i.e., the estimated AUCs being close to the true AUCs), while LMJM had the worst performance in both calibration and discrimination.

**Table 2.** Prediction results in Simulation II Scenario 1 in which random errors are generated from ALD with  $\tau = 0.25$ .

$t$	$\Delta t$	true $AUC_t^{\Delta t}$	QRJM ( $\tau = 0.25$ )		QRJM ( $\tau = 0.5$ )		LMJM	
			$S(t + \Delta t t)_{MSE}$	predicted $AUC_t^{\Delta t}$	$S(t + \Delta t t)_{MSE}$	predicted $AUC_t^{\Delta t}$	$S(t + \Delta t t)_{MSE}$	predicted $AUC_t^{\Delta t}$
0.25	0.25	0.823	0.006	0.815	0.137	0.814	0.244	0.782
	1	0.877	0.010	0.876	0.111	0.856	0.177	0.789
	2	0.918	0.012	0.905	0.083	0.894	0.126	0.795
	3	0.937	0.013	0.925	0.072	0.915	0.107	0.798
0.5	0.25	0.811	0.005	0.787	0.128	0.789	0.219	0.770
	1	0.869	0.013	0.843	0.127	0.843	0.221	0.782
	2	0.912	0.016	0.884	0.102	0.884	0.174	0.790
	3	0.929	0.017	0.910	0.091	0.903	0.153	0.794
0.75	0.25	0.783	0.005	0.767	0.110	0.766	0.189	0.759
	1	0.868	0.016	0.840	0.136	0.839	0.253	0.774
	2	0.906	0.020	0.879	0.117	0.880	0.218	0.784
	3	0.925	0.022	0.899	0.105	0.899	0.197	0.789

Bland-Altman plot,<sup>21</sup> a commonly used method to assess the agreement of two measurement methods, is applied to visually compare the predicted results against the true values. Web Figure 2 gives an intuitive comparison of the predicted results with the “gold standard” among different models. Plots from the true model (Web Figure 2.1) are horizontally spindle-shaped, suggesting that it is easier to predict a survival probability near 0 or 1 than the middle probability area near 0.5. Further, with an increase in  $\Delta t$ , there is more variation in the middle probability area, indicating that survival probability predictions for time points further into the future are less accurate than predictions for closer time points, as expected. Bland-Altman plots from the other models (QRJM with  $\tau = 0.5$ , Web Figure 2.2 and LMJM, Web Figure 2.3) display systematically biased patterns in predictions, which are consistent with the findings in Table 2. Simulation results for Scenario 3 are displayed in Web Table 3 and Web Figure 3, in which QRJM with  $\tau = 0.5$  performs comparably well as the LMJM (true model) when random errors are generated from standard normal distribution, suggesting the robustness of median regression model.

In summary, when data are simulated from ALD with specific skewness  $\tau$ , the best predictions of survival probability are obtained using the QRJM model with the exact quantile that generates the outcome data. Instead, LMJM results in systematically biased predictions when data are skewed and/or heavy-tailed. When random errors are normally distributed, predictions from QRJM with  $\tau = 0.5$  are comparable with those from the true LMJM.

## 4 Application

### 4.1 The Predictors of Huntington's Disease (PREDICT-HD) Study

The motivating PREDICT-HD study is an observational study that aims to identify the earliest signs of HD onset so that future HD drug trials can be targeted toward treatment that may slow the progression of the disease, or prevent it altogether.<sup>22</sup> Briefly, HD is known to be caused by the mutation of the first exon of the Huntington (HTT) gene, where expansion of the cytosine-adenine-guanine (CAG) is observed for HD patients.

The study recruited individuals from 33 medical centers in six countries (i.e., USA, Canada, Germany, Australia, Spain, and UK) starting from August, 2002. Qualified individuals were healthy pre-HD people without any symptoms of HD; i.e., who have not had a motor diagnosis of HD based on the Unified Huntington Disease Rating Scale (UHDRS) and had more than 35 HTT CAG repeats. Detailed inclusion and exclusion criteria can be found in Paulsen *et al.*<sup>23</sup> Baseline demographic information such as age, gender, education years, as well as clinical variables such as CAG repeat length and Beck Depression Inventory were recorded at enrollment. The data used in the current study contains 1078 individuals enrolled until July, 2014. Among those 1078 individuals, 64% are female and the mean age at baseline is 39.8 years (SD=10.39, range 18.1-83.7), education is 14.5 years (SD=2.60, range 8.0-20.0), and number of CAG repeats is 42.5 (SD=2.69, range 12-62). In this study, the time variable is defined as months since enrollment. The average follow-up time is 61.2 months (SD=39.6; range 0.12-144.0) and 959 (89%) individuals have data for at least two years. The survival event of interest is time to motor diagnosis of HD since enrollment, which is defined as having a diagnostic confidence level (DCL) score of 4 (the highest score).<sup>24</sup> During the study follow-up, 225 (21%) events (HD onsets) were observed.

We use total motor score (TMS) as the longitudinal outcome and consider the following QRJM for data analysis:

$$\begin{cases} y_i(t) = m_{it} + \varepsilon_{it} = \beta_0 + \beta_1 t + \beta_2 \text{age}_i + u_{i1} + u_{i2} t + \varepsilon_i(t), \varepsilon_i(t) \sim \text{ALD}(0, \sigma, \tau) \\ h(T_i | \mathcal{M}_{iT_i}; \gamma, \alpha) = \lambda_0(t) \exp(\gamma_1 \text{education}_i + \gamma_2 I_{\text{male}_i} + \alpha m_i(T_i)), \end{cases}$$

where  $y_i(t)$  represents the observed TMS score of individual  $i$  at visit time  $t$  and  $\text{age}_i$  is the baseline age. In the survival sub-model, we specify a piecewise constant baseline hazard function with three time intervals, i.e.,  $\lambda_0(t) = \sum_{k=1}^3 \lambda_k I_k(T_i)$ , where  $\lambda_k$  is the hazard rate for time interval  $[t_k, t_{k+1})$  and  $I_k(t) = 1$  if  $t \in [t_k, t_{k+1})$  and 0 otherwise. In addition to three intervals, we explored various numbers of pieces for the baseline hazard function and obtained very similar results (result not shown).

## 4.2 Dynamic Predictions of HD Risk

To evaluate the predictive accuracy of the model, we randomly split the 1078 study individuals into five subgroups with approximately equal sizes and conduct five-fold cross validation. Each of the five subgroups will be used as the validation data once to make out-of-sample predictions based on the inference result from the data with the validation samples left out. In each model fitting step, two chains are initiated with diverse initial values and the chains are considered to converge if the potential scale reduction factors (PSRF)<sup>25</sup> for all parameters are below 1.1. In the prediction step we choose different censoring time ( $t$ ) and prediction time window ( $\Delta t$ ) combinations. Predicted survival probabilities for all individuals are combined and used in the calculation of time-dependent AUC. We also calculate the AUCs from the LMJM and compare them with those from our QRJM model. Results are summarized in Table 3. There are several interesting “patterns” observed in Table 3. First, for a fixed censoring time  $t$ , time-dependent AUC usually increases with the increase in  $\Delta t$  (this is sensible in that HD-free probability decreases with time, thus it is easier for the model to differentiate those who will versus those who will not experience the event at later time points). However, with excessive long follow-up time and large prediction interval (e.g.  $t = 24$  and  $48$ ,  $\Delta t = 36$ ), most patients have low event-free probabilities and hence it is more difficult to differentiate them, leading to decreasing AUC. Second, for the same  $t + \Delta t$  value, we see an increasing trend in AUC from  $t = 12$  months to 24 months and 48 months (e.g.,  $t = 12$ ,  $\Delta t = 24$  v.s.  $t = 24$ ,  $\Delta t = 12$  or  $t = 24$ ,  $\Delta t = 36$  v.s.  $t = 48$ ,  $\Delta t = 12$ ). The improvement in predictive performance can be due to longer follow-up time with additional thus more informative longitudinal data. Third, compared with the LMJM, the QRJM has better prediction at some quantiles, but not all, because some quantiles of the outcome can be more informative in predicting future survival outcome than the conditional mean while some are not. There are also few cases where the QRJM has better predictions at all quantiles (e.g.  $t = 12$  at all  $\Delta t$ ) and vice versa (e.g.  $t = 24$ ,  $\Delta t = 12$ ).

We select three individuals (with IDs 12, 63, and 110) with different TMS trajectories (Figure (2.1)) to illustrate how our method provides subject-specific dynamic predictions of HD-free survival. For each individual, predictions are calculated and then updated based on increasing number of longitudinal measurements. In Figure (2.2), we have the longitudinal measures up to 24 months ( $t = 24$ ) for each individual, and predictions are made for  $\Delta t$  being 12, 24, 36, and 48 months respectively. We then increase the follow-up time to 36 and 48 months ( $t = 36$  and  $t = 48$ ) and update the predictions using the same  $\Delta t$  values. Updated prediction results are displayed in Figure (2.3) ( $t = 36$ ) and Figure (2.4) ( $t = 48$ ), respectively. The plots suggest that individuals with lower (better) and stable TMS have much higher HD-free probability (i.e., lower predicted risk of HD onset, individual 12 v.s. individual 62 and 110). With longer follow-up times and more longitudinal measurements, predicted HD-free probabilities

tend to be more accurate as indicated by the narrower point-wise 95% credible intervals ( $t = 48$  months v.s.  $t = 24$  months).

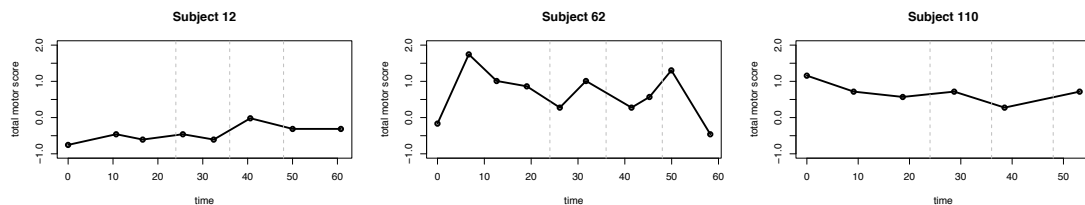
**Table 3.** PREDICT-HD data analysis: time-dependent AUC of the predictions of HD-free probability from QRJM and LMJM with TMS as the longitudinal biomarker.

$t$ (month)	$\Delta t$	AUC (QRJM)			AUC (LMJM)
		$\tau = 0.25$	$\tau = 0.50$	$\tau = 0.75$	
12	12	0.759	0.765	0.751	0.570
	24	0.802	0.791	0.773	0.695
	36	0.821	0.815	0.798	0.684
24	12	0.862	0.848	0.827	0.897
	24	0.858	0.850	0.831	0.853
	36	0.837	0.833	0.814	0.844
48	12	0.865	0.891	0.904	0.918
	24	0.879	0.877	0.881	0.879
	36	0.770	0.785	0.795	0.774

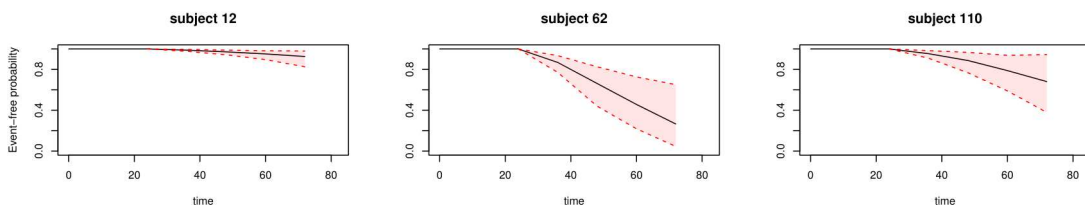
### 4.3 Inference Results for PREDICT-HD data

Finally, Table 4 presents the inference result from the QRJMs with  $\tau = 0.25, 0.50$ , and  $0.75$  and TMS as the longitudinal biomarker. In the longitudinal process, the coefficient for time quantifies the change rate of TMS at a fixed quantile. For example, Table 4 suggests that the time effect is  $0.019$  (95% CI:  $0.015, 0.023$ ) for TMS at  $\tau = 0.25$ , indicating one month increase in time is associated with  $0.019$  unit increase of 25% conditional quantile of TMS. The positive time coefficients for TMS at all quantiles indicate that TMS increases (deteriorates) over time, which is consistent with the loess curve in Figure 1 (upper panel). The magnitude of the time coefficients is similar at different quantiles of TMS, indicating comparable progression at those quantiles.

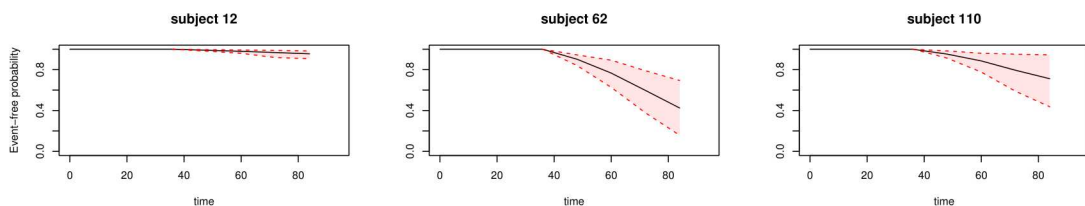
Table 4 also suggests that higher baseline age is associated with higher (worse) TMS and one year increase in baseline age is associated with  $0.004$  (95% CI:  $0.001, 0.008$ ) units increase in TMS at  $\tau = 0.5$ . In the time to HD onset process, across all three quantiles, education has protective effect on HD onset, i.e., those with more education years tend to have a lower risk of HD onset. Outcome TMS is strongly predictive of HD onset at all quantiles. Specifically, an unit increase in TMS increases the risk of HD onset by  $4.600$  ( $e^{1.526}$ , 95% CI  $3.747$ - $5.726$ ) times at  $\tau = 0.25$ , by  $3.669$  (95% CI  $3.152$ - $4.302$ ) times at  $\tau = 0.5$ , and by  $2.945$  (95% CI  $2.633$ - $2.294$ ) times at  $\tau = 0.75$ , respectively.



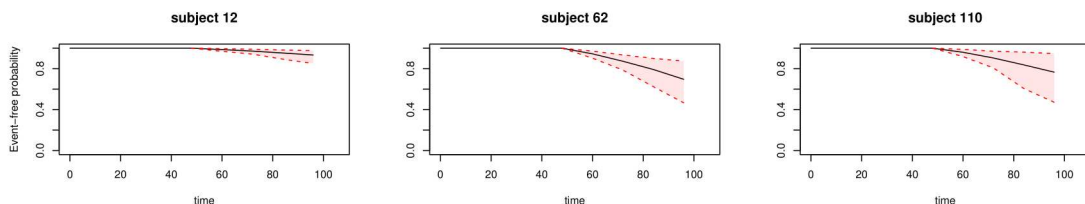
(2.1) Longitudinal trajectories of TMS for three selected individuals.



(2.2) Predictions based on censoring time  $t = 24$  months



(2.3) Predictions based on censoring time  $t = 36$  months



(2.4) Predictions based on censoring time  $t = 48$  months

**Figure 2.** PREDICT-HD data analysis: Longitudinal trajectories and dynamic predictions of HD-free probabilities with 95% pointwise credible interval from QRJM at  $\tau = 0.5$  for selected individuals.

## 5 Discussion

In our application of the linear-mixed joint model (LMJM, a linear mixed sub-model for the longitudinal process and a Cox sub-model for the survival process) to Huntington's Disease, there are two limitations.



**Table 4.** PREDICT-HD data analysis: Parameter estimation and 95% credible interval (in parenthesis) from QRJM at three different quantiles with TMS being the longitudinal biomarker.

	$\tau = 0.25$	$\tau = 0.50$	$\tau = 0.75$
For longitudinal TMS process			
Intercept	-0.760 (-0.903, -0.628)	-0.525 (-0.699, -0.359)	-0.249 (-0.469, -0.035)
Time (months)	0.019 (0.015, 0.023)	0.020 (0.016, 0.024)	0.022 (0.018, 0.026)
Age	0.004 (0.001, 0.008)	0.005 (0.001, 0.010)	0.006 (0.001, 0.012)
For time to HD onset process			
Education (years)	-0.083 (-0.115, -0.052)	-0.112 (-0.142, -0.082)	-0.128 (-0.157, -0.101)
Male	0.317 (-0.037, 0.654)	0.360 (-0.020, 0.708)	0.317 (-0.010, 0.647)
Association	1.526 (1.321, 1.745)	1.300 (1.148, 1.459)	1.080 (0.968, 1.192)

First, the normality assumption of the random errors in the linear mixed model (LMM) was not realistic, and no obvious transformation of the longitudinal outcome to produce residual normality was applicable. This limitation is confirmed by our simulation studies where the LMJM tends to provide biased estimates to model parameters when the longitudinal data are non-normal (skewed and /or heavy-tailed). Consequently, predictive accuracy of the LMJM is negatively impacted. Second, the LMJM only models the conditional mean of the outcome. However, in our (and other) clinical research application(s), it may be more clinically relevant to consider the tails of the outcome distribution, e.g., the upper tail of TMS is at higher risk of developing HD.

Our proposed quantile regression joint models (QRJM) uses a linear quantile mixed model (LQMM) for the longitudinal process, improves both inference and the ability to make accurate dynamic predictions. The quantile-based estimators are more robust against skewness in the data. Thus, our approach provides the flexibility to use median or quantile regression instead of mean regression when outliers and skewness are present in the longitudinal process. Moreover, the QRJM provides quantile-specific parameter estimates at a set of different quantiles and the researchers can choose the quantiles of interest and the corresponding inference results. The simulation studies and data application suggest that the QRJM not only inherits the good properties of an LMJM, but adds flexibility to the modeling procedure.

In this work, we develop a Bayesian algorithm to fit the proposed QRJM model and make dynamic predictions using the location-scale representation of the asymmetric Laplace distribution (ALD) for the longitudinal quantile regression. The Bayesian algorithm, which is straightforwardly implemented in JAGS software, uses a piecewise constant baseline hazard function in the survival sub-model. However, other functional forms can also be considered and the integration of the hazard function can be approximated using numerical integration such as Simpson's rule. In the real data application, we illustrate the flexibility of the QRJM and its advantages over the LMJM by jointly modeling the risk of

developing HD and total motor score (TMS, a commonly used early predictor of HD). The QRJM is able to provide more insight into the disease progression and the association between the two disease processes in terms of various quantile-based estimations and dynamic predictions.

The novel application of our proposed QRJM in making personalized dynamic predictions of survival probability finds practical importance in many clinical applications. Event prediction using commonly collected biomarkers can provide clinicians with continuously updated “disease progression” information potentially allowing them to make appropriately timed intervention decisions for each individual. Subject-specific dynamic predictions models play an important role in the ongoing transformation of traditional medicine to personalized care. While in traditional medicine, treatment selection is guided by population-based experience, i.e., the average treatment effect in the entire population, such “one-size-fits-all” approaches are criticized for resulting in low efficacy and high adverse drug reactions in many clinical studies. Utilizing the dynamic predictions based on the proposed QRJM framework, we obtain subject-specific predictions of event risk that would actually allow physicians to tailor medical treatment based on a specific patient profile.

In summary, the QRJM is a good alternative to the LMJM when either the normality assumption of the errors term is concerned or the conditional quantiles are more relevant to the research questions. When the interest lies in the prediction of future survival probability, the “best” quantile may be chosen based on the predictive accuracy criteria. Other model selection methods or methods, e.g., Bayesian model averaging, to incorporate multiple regression results from different quantiles into a single prediction solution can be a potential direction for future work. However, in our opinion, the choice of specific quantile(s) should be more clinically oriented rather than be a statistical task.

## Acknowledgements

Sheng Luo’s research was supported by the National Institute of Neurological Disorders and Stroke under Award Number R01NS091307. The authors acknowledge the Texas Advanced Computing Center (TACC) for providing high-performing computing resources. URL: <http://www.tacc.utexas.edu>.

## References

1. Self S and Pawitan Y. Modeling a marker of disease progression and onset of disease. In *AIDS Epidemiology*. Springer, 1992. pp. 231–255.
2. Tsiatis AA, Degruttola V and Wulfsohn MS. Modeling the relationship of survival to longitudinal data measured with error. applications to survival and cd4 counts in patients with aids. *J Am Stat Assoc* 1995; 90(429): 27–37.

3. Tsiatis AA and Davidian M. Joint modeling of longitudinal and time-to-event data: an overview. *Stat Sin* 2004; 14(3): 809–834.
4. Laird N and Ware J. Random-effects models for longitudinal data. *Biometrics* 1982; 38(4): 963–974.
5. Koenker R and Hallock K. Quantile regression: An introduction. *J Econ Perspect* 2001; 15(4): 43–56.
6. Koenker R. *Quantile Regression*. Cambridge University Press, 2005.
7. Koenker R and Basset G. Regression quantiles. *Econometrica* 1978; 46(1): 33–50.
8. Koenker R. Quantile regression for longitudinal data. *J Multivar Anal* 2004; 91(1): 74–89.
9. Geraci M and Bottai M. Quantile regression for longitudinal data using the asymmetric laplace distribution. *Biostatistics* 2007; 8(1): 140–154.
10. Geraci M and Bottai M. Linear quantile mixed models. *Stat Comput* 2014; 24(3): 461–479.
11. Kozumi H and Kobayashi G. Gibbs sampling methods for bayesian quantile regression. *J Stat Comput Simul* 2011; 81(11): 1565–1578.
12. Farcomeni A and Viviani S. Longitudinal quantile regression in the presence of informative dropout through longitudinal–survival joint modeling. *Stat Med* 2015; 34(7): 1199–1213.
13. Rizopoulos D. Dynamic predictions and prospective accuracy in joint models for longitudinal and time-to-event data. *Biometrics* 2011; 67(3): 819–829.
14. Taylor JMG, Park Y, Ankerst DP et al. Real-time individual predictions of prostate cancer recurrence using joint models. *Biometrics* 2013; 69(1): 206–213.
15. Koenker R and Machado JA. Goodness of fit and related inference processes for quantile regression. *J Am Stat Assoc* 1999; 94(448): 1296–1310.
16. Yu K and Moyeed RA. Bayesian quantile regression. *Statistics & Probability Letters* 2001; 54(4): 437–447.
17. Kotz S, Kozubowski T and Podgorski K. *The Laplace Distribution And Generalizations: A Revisit with Applications to Communications, Economics, Engineering, And Finance*. Springer Science & Business Media, 2012.
18. Rizopoulos D, Hatfield LA, Carlin BP et al. Combining dynamic predictions from joint models for longitudinal and time-to-event data using bayesian model averaging. *J Am Stat Assoc* 2014; 109(508): 1385–1397.
19. Plummer M. Jags: A program for analysis of bayesian graphical models using gibbs sampling. In *Proceedings of the 3rd international workshop on distributed statistical computing*, volume 124. Vienna, p. 125.
20. Heagerty PJ, Lumley T and Pepe MS. Time-dependent roc curves for censored survival data and a diagnostic marker. *Biometrics* 2000; 56(2): 337–344.
21. Bland JM and Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *The Lancet* 1986; 327(8476): 307–310.
22. Paulsen JS, Langbehn DR, Stout JC et al. Detection of huntington’s disease decades before diagnosis: the predict-hd study. *J Neurol Neurosurg Psychiatry* 2008; 79(8): 874–880.

23. Paulsen JS, Hayden M, Stout JC et al. Preparing for preventive clinical trials: the predict-hd study. *Arch Neurol* 2006; 63(6): 883–890.
24. Paulsen JS, Long JD, Ross CA et al. Prediction of manifest huntington’s disease with clinical and imaging measures: a prospective observational study. *The Lancet Neurology* 2014; 13(12): 1193–1201.
25. Brooks SP and Gelman A. General methods for monitoring convergence of iterative simulations. *J Comput Graph Stat* 1998; 7(4): 434–455.