Model Estimation and Dynamic Predictions for Subject Specific Event Probabilities in Joint Modeling Using Longitudinal Quantile Regression

Ming Yang

Contents

List of Tables

List of Figures

1 Background

1.1 Literature Review

1.1.1 General Background of Joint Modeling

Longitudinal studies are ubiquitous in biostatistics context. For example, in randomized clinical trials (RCT) patients are randomly allocated into different treatment arms and are then followed over time to collect outcome(s) and risk factors. Repeated measurements will then be produced from this follow-up mechanism. One of the important features of longitudinal data is that the repeated measurements from the same subject are more "similar" to each other compared to those measures from different subjects, i.e. within subject measures tend to be intercorrelated. In statistical analyses, this feature requires special techniques to handle the correlation for valid scientific inference. There are mainly three methods for analyzing longitudinal data: marginal model, transition model and random effects model (Diggle et al., 2002). Estimations of the regression coefficients have different interpretations and the choice of a model depends on study objectives, the source of correlation as well as the capacity of the model. This thesis work will focus on applying random effects models to longitudinal data. A model that contains both random effects and fixed effects is called mixed effects model. The mixed effects model methodology is a statistical tool that is used across a wide variety of disciplines including biostatistics. Currently, mixed effects models are especially popular in research involving repeated measurements (Laird and Ware, 1982) or observations from multilevel (or hierarchical) structure where the correlation between observations is not negligible.

In many clinical trials and medical studies, time-to-event data are commonly generated along with the longitudinal measurements. Often, the outcome of interest in survival data, such as disease recurrence, possible drop-outs, or death, is correlated with the longitudinal measurements. For example, HIV patients with decreasing CD4 cell count are more likely to die or prostate cancer patients with elevated prostate specific antigen (PSA) are more susceptible to prostate cancer recurrence. Simply ignoring the correlation and fitting two models separately will lead to loss of information and misleading results. The joint model (JM) method for longitudinal and survival data was first proposed by Tsiatis et al. (1995) and Faucett and Thomas (1996) to handle this issue and to obtain unbiased estimators. JM is well studied in recent years, for examples see Henderson et al. (2000), Wang and Taylor (2001), and Xu and Zeger (2001). Guo and Carlin (2004) developed a fully Bayesian method to fit the JM using MCMC methods and implemented them in WinBUGS software. For more details, see Yu et al. (2004) as a good review of the JM methodology. Many

extensions have also been developed for JM, including considering multiple longitudinal outcomes (Brown et al., 2005; Rizopoulos and Ghosh, 2011), incorporating multiple failure times (Elashoff et al., 2008).

1.1.2 Longitudinal Quantile Regression in the Setting of Joint Modeling

In most of the JM related works the longitudinal part is modeled using the linear mixed model (LMM), which is a widely used application of the mixed effects method. In brief, an LMM assumes the expected value of the longitudinal outcome is a linear function of the covariates and repeated observations from the same subject share a same unobserved latent variable, i.e. random effect, to account for the correlation between them. When conditional on random effects, observations from the same subject are treated as independent. In addition, traditional LMM also assumes the distribution of unobserved random error is Gaussian.

Our concern for the widely used LMM is that in many circumstances the normality assumption of the error term cannot be satisfied (even after trying various transformations). A commonly encountered situation is when outliers exist or when the outcome is skewed. In these situations, LMM is not appropriate to use directly. In other cases, the conditional mean of the longitudinal outcome may not be the primary interest and researchers may be more interested in the covariates effect on the lower/upper quantiles of the outcome. For example, the treatment effect for patients with higher blood pressure is of greater importance to us because they are at higher risk of having strokes or developing heart failures. Instead of trying to fix the limitations of LMM, quantile regression is an alternative method that provides a single solution to all above issues with LMM. There are several advantages of quantile regression over the ubiquitous mean regression (or linear regression) model. To list a few, quantile regression provides a much more comprehensive and focused insight into the association between the variables by studying the conditional quantile functions of the outcome, which may not be observed by looking only at conditional mean of the outcome (Koenker, 2005). In quantile regression, the regression coefficients ($\boldsymbol{\beta}$) are functions of the quantile (τ), thus the estimated values of $\boldsymbol{\beta}$ vary according to different quantiles. As a results quantile regression provides a way to study the heterogeneity of the outcome that is associated with the covariates (Koenker, 2005). Moreover, quantile regression is more robust against outcome outliers compared with the mean regression, which is an immediate extension from the property of quantiles.

Quantile regression is becoming more and more popular in the statistical community. Koenker and Bassett Jr (1978) introduced a method for estimating the conditional quantiles. As an introductory material, Koenker and Hallock (2001) briefly covers the fundamentals of quantile regression, parameter estimation techniques,

inference, asymptotic theory, etc., Koenker's 2005 book provides a comprehensive and deeper introduction to quantile regression related topics (Koenker, 2005). Yu and Moyeed (2001) introduced the idea of Bayesian quantile regression by modeling the error term in the model using asymmetric Laplace distribution (ALD). Much work has been done to extend the quantile regression method to accommodate longitudinal data. Jung (1996) developed a quasi-likehood method for median regression model for longitudinal data. Geraci and Bottai (2007) proposed to fit the quantile regression for longitudinal data based on ALD and the estimation is made by using a Monte Carlo EM algorithm. Later on, Liu and Bottai (2009) followed the idea of Geraci and Bottai (2007) and extended the model from random intercept to including random slope as well. The study of longitudinal data using quantile regression has become popular in recent years. Fu and Wang (2012) proposed a working correlation model for quantile regression for longitudinal data. An induced smoothing method was used to make the inference of the estimators. Fully Bayesian techniques and Gibbs sampling algorithm become possible when the error term is decomposed as the mixture of normal and exponential random variables for the quantile linear mixed model; see Kozumi and Kobayashi (2011) and Luo et al. (2012) for applications. The fully Bayesian method is appealing because it is easy to implement, easy make inferences, the uncertainty of the unknowns is taken into account, and it is flexible in the distribution of random effects. The detailed background about Bayesian quantile linear mixed model will be provided in Section 3.1.2.

1.1.3 Subject Specific Dynamic Predictions Based on Joint Modeling

In recent years, another extension of the JM that attracts increasing attention is to make subject specific predictions for longitudinal or survival outcomes based on patient information at hand. In clinical settings, as we monitor the health of a patient over time under a joint modeling frame work, time-varying measurements can be used to derive other useful summary indicators such as probability of events. Thus JM provides a vital tool in predicting future health outcomes for the patients at risk.

There are several applications of this prediction idea. Yu et al. (2008) used a JM framework to study the longitudinal measures of PSA in predicting the probabilities of recurrence of prostate cancer up to four years in the future. In their work, the longitudinal part was modeled using a nonlinear hierarchical mixed model and Cox proportional hazards model with time-dependent covariates was used to model the time to clinical recurrences. Both the value of longitudinal outcome and probability of recurrence were predicted. Proust-Lima and Taylor (2009) also worked on the PSA and prostate cancer problem and they used the joint

latent class model (JLCM) (Lin et al., 2002) to build a dynamic prognostic tool for predicting the recurrence of prostate cancer, in which they used the maximum likelihood estimate method. As another example, Rizopoulos (2011) illustrated how to make survival probabilities predictions under the JM framework using the frequentist method. The application is demonstrated using the famous AIDS CD4 cell count data. Another important component for all of the prediction works is to validate the accuracy of the predicted results, which is discussed in all of above studies. There are different aspects to consider for assessing the predictive performance of the model. Rizopoulos (2011) used the receiver operating characteristic (ROC) curve based an approach to assess how well the proposed model can differentiate patients who will have events from those who will not. While Proust-Lima and Taylor (2009) computed the absolute error of prediction (EP) curves showing weighted average absolute error of prediction (WAEP) over three years and the EP at one- and three- year horizons. And in Taylor et al. (2013), a simple graphical method was used to assess the predicative accuracy. More technique details will be given in Section 3.3 about the JM estimation, prediction and results validation that will be used in this thesis work.

1.2 Public Health Significance

This thesis work contributes to the public health field in the following ways. Studying the low or high tail of the longitudinal outcome can be of greater interest and more meaningful compared with focusing on the population mean. Specific examples of preference for studying conditional quantiles over the conditional mean of the outcomes are cardiovascular studies that focus on the effect of interventions to reduce the blood pressure for hypertensive patients (upper tail) who are clinically at greater risk of developing heart diseases and study of low birth weight infants (lower tail) (Koenker and Hallock, 2001). Under such conditions, our method provides a vital complement to the traditional linear regression method. Due to the flexibility of quantile regression in modeling the longitudinal outcome, i.e. regression quantile can be chosen depending on specific research interest, we would be able to gain a much better insight on the relationship between the outcome and covariates. Under the JM framework, the regression parameters in the survival model are also functions of quantile, which means our model will provide the quantile specific association between the longitudinal outcome and the event probability.

Another, and more important, utilization of our model is to make "accurate" subject specific predictions of future event probabilities, which can be of great importance in clinical practice. The parameter estimations of the predictive model will be made based on a relatively large sample (i.e. the training set), when given a

new patient with his or her baseline characteristics as well as the historical biomarker records, the model is able to produce subject specific prediction for this patient. This would allow us to tailor medical treatment specifically for that patient in order to lower his or her event probabilities in the future. This idea fits into the big concept of "personalized medicine", which aims to provide the right patient with the right drug at the right time (FDA, 2013). Traditional health care interventions prescribed to patients are designed based on its average effect on the population, but these interventions may work perfectly on some patients but not on others. Thus, individualized treatment should be more effective. The practice of customized treatment does already exist in treating tumors where doctors prescribe drugs based on the tumor's growth and some specific gene mutations of the patient, but there are many more potential applications of personalized medicine for patients suffering with other diseases. Thus, as Precision Medicine Initiative strives for, our method of making subject specific predictions of event probabilities in the future provides an important tool to physicians in making individualized decisions in order to achieve better clinical outcomes.

1.3 Research Questions and Specific Aims

The joint model of longitudinal data, using quantile regression and survival data is little studied. To our knowledge, Farcomeni and Viviani (2014) is the first work that extended classical JM to incorporate a quantile regression model in the longitudinal process. In their paper the parameter estimations are obtained using the Monte Carlo expectation and maximization (MCEM) method. Also, there is no work extending the subject specific dynamic predictions method to the longitudinal quantile regression based JM framework. We aim to extend current research and fill those gaps.

We have the following aims:

- To develop a fully Bayesian method for estimating the model parameters in the proposed JM, based on which a subject specific dynamic predictions method for survival events will be developed;
- To extend our new Bayesian JM method in Aim 1 to study the recurrent events data and develop a new method for parameter estimation in the JM of longitudinal and recurrent event data;
- To develop a method for subject specific dynamic predictions of recurrent events based on our results from the first and second aims.

To demonstrate the application of our proposed methods, we will be using the data from The Antihypertensive and Lipid-Lowering Treatment to Prevent Heart Attack Trial (ALLHAT) (Davis et al., 1996), which

was the largest antihypertensive treatment trial and the second largest lipid-lowering trial ever conducted. We will mainly focus on the antihypertensive component of the trail in studying the effects of risk factors as well as to make dynamic predictions of event occurrences for subjects at the higher tail of the longitudinal outcome (i.e. the blood pressure). More information about the ALLHAT study is given in Section 2.

2 Data Description

Hypertension induced heart failure (HF) is a major cause of hospitalization for American seninors. There are over 5.8 million people in the US suffering with HF and more than 670000 new cases joining this group every year. Hypertension treatments that have been studied include thiazide-type diuretics and angiotensin converting enzyme inhibitors (ACEIs), both of which have been shown to be effective in reducing the incidence of HF by treating hypertension.

As new treatment agents were developed, ALLHAT compared the effects of different treatment methods on fatal coronary heart disease (CHD) or non-fatal myocardial infarction (MI) among high-risk hypertensive patients. Four treatment methods under comparison were thiazide-type diuretic (chlorthalidone), ACEI (lisinopril), calcium channel-blocker (CCB; amlodipine), and α-blocker (doxazosin), in which the later three are the newer antihypertensive agents whose effects were not previously studied. Under the multi-center, randomized, double-blind, active-controlled design mechanism, in the ALLHAT study a total of 42448 participants were randomized from 625 sites in the United States, Canada, Puerto Rico, and the US Virgin Islands (Grimm et al., 2001). The study included a large cohort of African Americans (36% of total study population) and a relatively large proportion of Hispanic patients (19%). Other major baseline characteristics of the study cohort include almost equal proportion of each gender (46.8% women), average age of 67 years (with 35% aged $\geq$70 years), high proportion of patients with diabetes (36%), 47% of the cohort with existing cardiovascular disease and 22% are smokers. More detailed baseline characteristics of participants in the ALLHAT study can be found in Grimm et al. (2001).

The full-scale ALLHAT study was started in fall of 1994 and the participants were then followed actively until March 31, 2002. Subsequently, the participants were then followed through 2006 in a national extended follow-up (Piller et al., 2011). As a result, the average of total follow-up time is 8.9 years: 4.9 years (3.2 years in the doxazaosin v.s. chlorthalidone comparison part) active follow-up time plus 4 years of extended follow-up. The primary outcome was cardiovascular mortality and the secondary outcomes were mortality,

stroke, CHD, HF, cardiovascular disease, and end-stage renal disease. Cushman et al. (2012) has presented the mortality and morbidity results during and after the trial. During the active follow-up stage, outcomes as well as other clinical information like blood pressure are all available, however, during the post-trail follow-up no information is available for medications or blood pressure. In this thesis work, we will be using this ALLHAT dataset as the base population to make model estimations, based on which the subject specific dynamic predictions will be conducted either for patients from this study cohort or patients outside of the study cohort but with similar baseline characteristics.

3 Statistical Methods

Three major statistical methods that will be used, including Bayesian quantile regression, JM uses longitudinal quantile regression and dynamic prediction method for the probabilities of future events based on JM of longitudinal and survival data.

3.1 Bayesian Quantile Regression

3.1.1 Quantile Regression and Asymmetric Laplace Distribution

Let Y be a real valued random variable with cumulative distribution function $F_Y(y) = P(Y \leq y)$. By definition, the τth quantile of Y, where $\tau \in [0, 1]$, is given by

$$Q_Y(\tau) = F_Y^{-1}(\tau) = \inf \{y : F_Y(y) \geq \tau\} \tag{1}$$

In contrast to mean regression (or linear regression), quantile regression models the conditional quantile of the outcome Y given a set of covariates, which is defined as

$$Q_{Y|\boldsymbol{X}}(\tau) = \boldsymbol{X}^\top \boldsymbol{\beta}_\tau \tag{2}$$

Given the data sample, the estimates of the regression coefficients at τth quantile can be obtained by solving

$$\hat{\boldsymbol{\beta}}_\tau = \underset{\boldsymbol{\beta} \in \mathbb{R}^p}{\arg\min} \sum_{i=1}^{n} \left[\rho_\tau (Y_i - \boldsymbol{X}_i^\top \boldsymbol{\beta}) \right], \tag{3}$$

where the loss function $\rho_\tau(\cdot)$ is defined as $\rho_\tau(Y) = Y(\tau - I(Y < 0))$.

However, there is no direct solution to (3), rather the minimization problem can be reformulated as a linear programming problem, where simplex methods or interior point methods can be applied to solve for the estimates (Koenker, 2005). The minimization problem can also be rephrased as a maximum-likelihood problem by using the asymmetrical Laplace distribution (ALD). (Koenker and Machado, 1999; Yu and Moyeed, 2001).

Suppose a random variable Y follows $\text{ALD}(\mu, \sigma, \tau)$, then the probability density function of Y can be written as

$$f(Y|\mu, \sigma, \tau) = \frac{\tau(1-\tau)}{\sigma} \exp\left[-\rho_\tau\left(\frac{Y-\mu}{\sigma}\right)\right], \tag{4}$$

where $\mu \in (-\infty, \infty)$ is the location parameter, σ is the scale parameter and $\tau \in (0, 1)$ is the skewness parameter. Thus, in a standard linear model

$$Y_i = \boldsymbol{X}_i{}^T\boldsymbol{\beta} + \varepsilon_i, \tag{5}$$

if we assume the random error $\varepsilon_i \sim \text{ALD}(0, \sigma, \tau)$, then $Y_i|\boldsymbol{X}_i \sim \text{ALD}(\boldsymbol{X}_i{}^\top\boldsymbol{\beta}, \sigma, \tau)$, that is the likelihood function can be written as

$$L(\boldsymbol{\beta}, \sigma; \boldsymbol{Y}, \tau) \propto \frac{1}{\sigma^n} \exp\left[-\sum_{i=1}^{n} \rho_\tau\left(\frac{Y_i - \boldsymbol{X}_i{}^\top\boldsymbol{\beta}}{\sigma}\right)\right]. \tag{6}$$

If we treat σ in (6) as nuisance then the maximization of (6) with respect to $\boldsymbol{\beta}$ is exactly the same as that in (3).

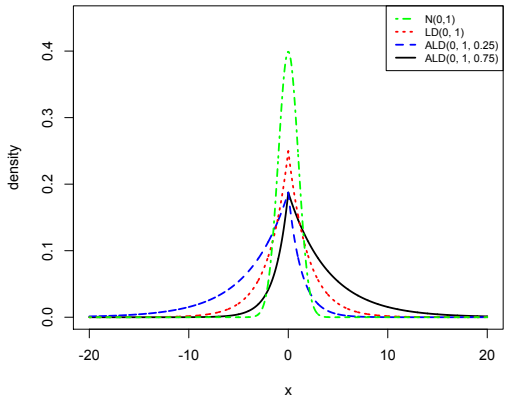

Figure 1: Graphical visualization of ALD versus Laplace distribution (LD) and standard normal distribution.

3.1.2 Bayesian Linear Quantile Mixed Model

As a natural extension to linear quantile regression, when working on longitudinal data, linear quantile mixed model (LQMM) is defined as

$$Q_{Y_{ij}|\boldsymbol{X}_{ij},\boldsymbol{Z}_{ij}}(\tau) = \boldsymbol{X}_{ij}^{\top}\boldsymbol{\beta} + \boldsymbol{Z}_{ij}^{\top}\boldsymbol{u}_i,\ i = 1, \cdots, N;\ j = 1, \cdots, n_i. \tag{7}$$

where Y_{ij} is the response variable for subject i at time j, $\boldsymbol{X}_{ij}$ is the $p-$dimensional fixed effects covariates and $\boldsymbol{\beta}_{\tau}$ is the corresponding $p \times 1$ vector of fixed effects, while $\boldsymbol{Z}_{ij}$ is the $k-$dimensional random effects covariates and $\boldsymbol{u}_i$ is the corresponding $k \times 1$ vector of random effects for subject i. Under the assumption that the random error follows ALD$(0, \sigma, \tau)$, conditional on the random effects $\boldsymbol{u}_i$, Y_{ij}'s are independently and identically distributed as ALD$(\boldsymbol{X}_{ij}^{\top}\boldsymbol{\beta} + \boldsymbol{Z}_{ij}^{\top}\boldsymbol{u}_i, \sigma, \tau)$:

$$f(Y_{ij}|\boldsymbol{\beta}, \boldsymbol{u}_i, \sigma) = \frac{\tau(1-\tau)}{\sigma} \exp\left[-\rho_{\tau}\left(\frac{Y_{ij} - \boldsymbol{X}_{ij}^{\top}\boldsymbol{\beta} - \boldsymbol{Z}_{ij}^{\top}\boldsymbol{u}_i}{\sigma}\right)\right] \tag{8}$$

To develop a Gibbs sampler for model (8), we need to assume a location-scale mixture representation of the ALD (Kotz et al., 2001). Under such assumption the random error is represented as $\varepsilon_{ij} = \kappa_1 e_{ij} + \kappa_2\sqrt{\sigma e_{ij}}v_{ij}$, where

$$\kappa_1 = \frac{1-2\tau}{\tau(1-\tau)},\ \text{and}\ \kappa_2^2 = \frac{2}{\tau(1-\tau)},$$

and

$$v_{ij} \sim N(0,1), \text{ and } e_{ij} \sim \exp(\sigma).$$

As a result, the linear mixed model is then reparameterized as

$$Y_{ij} = \boldsymbol{X}_{ij}^{\top}\boldsymbol{\beta} + \boldsymbol{Z}_{ij}^{\top}\boldsymbol{u}_i + \kappa_1 e_{ij} + \kappa_2\sqrt{\sigma e_{ij}}v_{ij}, \tag{9}$$

or equivalently

$$f(Y_{ij}|\boldsymbol{\beta}, \boldsymbol{u}_i, e_{ij}, \sigma) = \frac{1}{\sqrt{2\pi\kappa_2^2\sigma e_{ij}}} \exp\left[-\frac{1}{2\kappa_2^2\sigma e_{ij}}(Y_{ij} - \boldsymbol{X}_{ij}^{\top}\boldsymbol{\beta} - \boldsymbol{Z}_{ij}^{\top}\boldsymbol{u}_i - \kappa_1 e_{ij})^2\right]. \tag{10}$$

Following (9), a fully specified Bayesian model would include follows:

$$\begin{aligned}
&\boldsymbol{v} \sim \prod_{i=1}^{N}\prod_{j=1}^{n_i} \exp\left(-\frac{v_{ij}^2}{2}\right),\\
&\boldsymbol{e} \sim \prod_{i=1}^{N}\prod_{j=1}^{n_i} \frac{1}{\sigma}\exp\left(-\frac{e_{ij}}{\sigma}\right),\\
&\boldsymbol{\beta} \sim MVN_p(\boldsymbol{0}, \boldsymbol{\Sigma}),\\
&\boldsymbol{u}_i|\eta \sim MVN_p(\boldsymbol{0}, \eta^2\boldsymbol{I}),\\
&\sigma \sim IG(a_0, b_0),\\
&\eta \sim IG(c_0, d_0).
\end{aligned}$$

3.2 Longitudinal Quantile Regression and Joint Modeling

Following Farcomeni and Viviani (2014), we extend the traditional joint modeling (JM) of longitudinal and survival data by using the longitudinal quantile mixed model in place of the linear mixed model.

In the time-to-event data let $T_i = min(T_i^*, C_i)$ be the observed event time for subject i ($i = 1, \cdots, n$), where T_i^* is the true underlying event time and C_i is the censoring time. Let Δ_i be the event indicator and define it as $\Delta_i = I(T_i^* < C_i)$, where $I(\cdot)$ is the indicator function. If $\Delta_i = 1$, i.e. $T_i^* < C_i$, we say an event is observed during the study period; in contrast, when $\Delta_i = 0$ there is no event observed until the end of the study or when the patient is lost follow-up (i.e. censored).

While in the longitudinal part, let Y_{it} be the continuous longitudinal outcome for subject i ($i = 1, \cdots, n$) measured at time t ($t = 1, \cdots, n_i$). Note that we can only observe Y_{it} when $t \leq T_i$, so the longitudinal outcome for subject i can be written as $\boldsymbol{Y}_i = \{Y_{it} : t \leq T_i\}$.

There is also a set of covariates in the model. In the longitudinal model, let $\boldsymbol{X}_{it}$ and $\boldsymbol{H}_{it}$ be the fixed effects covariates that are associated with the outcome and $\boldsymbol{Z}_{it}$ be the covariates associated with $k-$dimensional random effects $\boldsymbol{u}_i$; in the time-to-event model, we have $\boldsymbol{W}_i$ as the fixed effects covariates that are only associated with event time (not longitudinal outcome) and this model shares the same fixed effects covariates $\boldsymbol{H}_{it}$ and random effects covariates $\boldsymbol{Z}_{it}$ with the longitudinal model. Thus the two models are related by sharing some of the fixed and random variables, and the degree of associations from those two sources of measurements (observed and unobserved) are measured by another two parameters α_1 and α_2, respectively.

The proposed JM as described above can be written as a set of two models:

$$\begin{cases} Y_{it} = \boldsymbol{X}_{it}^\top \boldsymbol{\beta} + \boldsymbol{H}_{it}^\top \boldsymbol{\delta} + \boldsymbol{Z}_{it}^\top \boldsymbol{u}_i + \varepsilon_{it}, \varepsilon_{it} \sim ALD(0, \sigma, \tau) \\ h(T_i | \mathcal{T}_{iT_i}, \boldsymbol{W}_i; \boldsymbol{\gamma}, \alpha_1, \alpha_2) = h_0(T_i) \exp(\boldsymbol{W}_i^\top \boldsymbol{\gamma} + \alpha_1 \boldsymbol{H}_{iT_i}^\top \boldsymbol{\delta} + \alpha_2 \boldsymbol{Z}_{iT_i}^\top \boldsymbol{u}_i) \end{cases} \tag{11}$$

where the first equation is the linear quantile mixed model discussed in Section 3.1.2 and the second equation takes the format of Cox proportional hazards model where $h_0(\cdot)$ is the baseline hazard function.

Individual heterogeneity is captured by the term $\boldsymbol{Z}_{it}^\top \boldsymbol{u}_i$ in the model, which is the deviation of subject i from the population average. Note that the posterior estimates of the random effects can be used to draw subject specific predictions of future event probabilities or longitudinal outcome when the subject is from the study population that we used to fit the model.

Also note that in quantile regression, the parameter estimates are functions of the quantile. This is also true in the proposed JM. That is, parameters in the survival models like $\boldsymbol{\gamma}$ also vary according to different values of τ. Depending on research aims, different strategies may be taken to utilize the flexibility of the model. For example, if we are interested in the complete distribution of parameter estimates as a function of quantile, we can just run the model through the range of possible quantiles, collect and compare the results from different quantiles of the outcome. Less varying values of the estimate indicates a relatively stable effect from the corresponding covariate on the outcomes. In contrast, if the interest only lies in investigating the effect of lower or higher quantile for the longitudinal outcome on the survival probabilities as discussed in Section 1.1.2, we may just fix the quantile to a specific value and conduct the analysis.

3.2.1 The Survival Model

As the details for the longitudinal model has been discussed previously in Section 3.1.2, here we only focus on the survival component of the JM.

For subject i, the complete survival likelihood can be written as:

$$\begin{aligned} f(T_i, \Delta_i|\boldsymbol{u_i}) &= f(T_i|\mathcal{T}_{iT_i}, \boldsymbol{W}_i)^{\Delta_i} S(T_i|\mathcal{T}_{iT_i}, \boldsymbol{W}_i)^{1-\Delta_i} \\ &= h(T_i|\mathcal{T}_{iT_i}, \boldsymbol{W}_i)^{\Delta_i} S(T_i|\mathcal{T}_{iT_i}, \boldsymbol{W}_i)^{1-\Delta_i}, \end{aligned} \tag{12}$$

where $h(T_i|\mathcal{T}_{iT_i}, \boldsymbol{W}_i)$ is given in (11) and $S(\cdot)$ is the survival function, i.e.

$$S(T_i|\mathcal{T}_{iT_i}, \boldsymbol{W}_i) = \exp\left\{ - \int_0^{T_i} h_0(s) \exp(\boldsymbol{W}_i^\top \boldsymbol{\gamma} + \alpha_1 \boldsymbol{H}_{is}^\top \boldsymbol{\delta} + \alpha_2 \boldsymbol{Z}_{is}^\top \boldsymbol{u}_i) ds \right\}. \tag{13}$$

For the baseline hazard $h_0(s)$, parametric form like Weibull model can be used or it can be left unspecified. However, the choice of baseline hazard is not a main focus of the this project.

3.2.2 Complete Likelihood Function and Bayesian Inference

The complete joint likelihood for longitudinal and survival outcomes, assuming random effects are known, for the ith subject is given by

$$L(\boldsymbol{\theta}; T_i, \Delta_i, \boldsymbol{Y}_i, \boldsymbol{u}_i) = f(\boldsymbol{Y}_i|\boldsymbol{u}_i) f(T_i, \Delta_i|\boldsymbol{u}_i) f(\boldsymbol{u}_i|\boldsymbol{\Sigma}), \tag{14}$$

where vector $\boldsymbol{\theta}$ represents a collection of all the parameters used in each distribution function in (14), $f(T_i, \Delta_i|\boldsymbol{u}_i)$ is given in (12) and

$$f(\boldsymbol{Y}_i|\boldsymbol{u}_i) = \prod_{t=1}^{n_i} f(Y_{it}|\boldsymbol{u}_i),$$

in which $f(Y_{it}|\boldsymbol{u}_i)$ has the format of (10).

As proposed in (Farcomeni and Viviani, 2014), parameter estimation can be done using Monte Carlo EM (MCEM) algorithm, where they treated the random effects as the missing data. In the EM algorithm, the conditional expectation of the complete log likelihood with respect to posterior distribution of random effects

is approximated using the Monte Carlo method by sampling from posterior distribution. The maximization step is then conducted to find the maximum likelihood estimation (MLE) of the parameters based on the conditional expectation. In contrast to their method, to avoid the complexity in derivation of the expectation and maximization functions as well as obtaining the standard error of the estimates, we take advantage of the location-scale mixture representation of the ALD and propose a fully Bayesian inference approach for the parameters by using the Markov chain Monte Carlo (MCMC) method. Specifically, given the complete likelihood in (14), by choosing appropriate priors for the parameters, according to the Bayes theorem the posterior distribution is given by

$$f(\boldsymbol{\theta}|\boldsymbol{T}, \boldsymbol{\Delta}, \boldsymbol{Y}, \boldsymbol{u}) \propto \prod_{i=1}^{n} f(T_i, \Delta_i, \boldsymbol{Y}_i, \boldsymbol{u}_i; \boldsymbol{\theta}) f(\boldsymbol{\theta}) \tag{15}$$

where $\boldsymbol{T} = (T_1, T_2, \cdots, T_n)$, $\boldsymbol{Y} = (\boldsymbol{Y}_1, \boldsymbol{Y}_2, \cdots, \boldsymbol{Y}_n)$, $\boldsymbol{\Delta} = (\Delta_1, \Delta_2, \cdots, \Delta_n)$, $\boldsymbol{u} = (\boldsymbol{u}_1, \boldsymbol{u}_2, \cdots, \boldsymbol{u}_n)$, and $f(\boldsymbol{\theta})$ is the product of the prior distributions:

$$f(\boldsymbol{\theta}) = \pi(\boldsymbol{\beta})\pi(\boldsymbol{\delta})\pi(\boldsymbol{\gamma})\pi(\alpha_1)\pi(\alpha_2)\pi(\sigma)\pi(\boldsymbol{\Sigma})$$

Where $\boldsymbol{\Sigma}$ is the covariance matrix of the random effects distribution. Similarly as in Section 3.1.2, we may choose the following priors for the parameters

$$\boldsymbol{\beta} \sim MVN_p(\boldsymbol{b}_0, \boldsymbol{B_0}),$$

$$\boldsymbol{\delta} \sim MVN_k(\boldsymbol{d}_0, \boldsymbol{D_0}),$$

$$\boldsymbol{\gamma} \sim MVN_k(\boldsymbol{g}_0, \boldsymbol{G_0}),$$

$$\alpha_1, \alpha_2 \sim N(a_0, \sigma_a),$$

$$\sigma \sim IG(s_0, s_1).$$

in which all the hyperparameters will be chosen so that the priors tend to be "flat" or non-informative. For the covariance matrix of the random effects, i.e. $\boldsymbol{\Sigma}$, we use Cholesky decomposition representation for the prior. For example, a 2×2 covariance matrix can be decomposed as follows:

$$\boldsymbol{\Sigma} = \begin{bmatrix} w_{11}^2 & w_{21}w_{11} \\ w_{21}w_{11} & w_{22}^2 + w_{21}^2 \end{bmatrix}, \tag{16}$$

priors for w's then can be assigned. For example:

$$w_{ii} \sim unif(a, b),$$

$$w_{ij} \sim N(\mu_w, \sigma_w), \text{ for } i \neq j.$$

The fully Bayesian inference will be implemented using the JAGS software (Plummer et al., 2003), in which, for the survival part, the so-called "zero-trick" will be used as the survival likelihood is not included in the standard distribution list in JAGS. The JAGS model file will be attached in Appendix.

3.3 Dynamic Predictions and Validation

This section focuses on the methodology of making subject specific predictions of future survival probabilities, which is realized by calculating the expected values of future survival probabilities. The accuracy of the predicted values will be validated using a ROC based approach. Note that all the parameter below is quantile specific and for the sake of simplicity in notation, we just omit all the quantile suffix, i.e. $\boldsymbol{\theta}$ represents $\boldsymbol{\theta}_\tau$.

3.3.1 Predicting Survival Probabilities

Upon fitting the JM based on a reference population consists of n subjects, we are then interested in making predictions of survival probabilities for a subject given a set of his or her historical biomarker measurements, which is denoted as $\mathcal{Y}_i(t) = \{Y_i(s), 0 \leq s \leq t\}$, and baseline covariates. An implication of JM is that up to time t, until when the longitudinal measurements are available, the subject must be alive or free of event occurrence as $Y_i(t)$ serves as the time-dependent covariate in the survival model. Thus what we are really interested in is the survival probability up to time $m > t$ given the survival up to time t, i.e.,

$$p_i(m|t) = Pr(T_i^* \geq m | T_i^* > t, \mathcal{Y}_i(t), \mathcal{D}_n; \boldsymbol{\theta}), \tag{17}$$

where $\mathcal{D}_n = \{T_i, \Delta_i, \boldsymbol{Y}_i, i = 1, \cdots, n\}$ denotes data from the reference population of size n, based on which we fit our JM.

Equation (17) can be furtherer developed as

$$\begin{aligned} Pr(T_i^* \geq m | T_i^* > t, \mathcal{Y}_i(t), \mathcal{D}_n; \boldsymbol{\theta}) &= \int Pr(T_i^* \geq m | T_i^* > t, \mathcal{Y}_i(t), \mathcal{D}_n, u_i; \boldsymbol{\theta}) \times \\ & \quad Pr(u_i | T_i^* > t, \mathcal{Y}_i(t), \mathcal{D}_n; \boldsymbol{\theta}) du_i \\ &= \int Pr(T_i^* \geq m | T_i^* > t, u_i; \boldsymbol{\theta})) Pr(u_i | T_i^* > t, \mathcal{Y}_i(t); \boldsymbol{\theta}) du_i \\ &= \int \frac{S_i[m | \mathcal{M}_i(m, u_i, \boldsymbol{\theta}); \boldsymbol{\theta})]}{S_i[t | \mathcal{M}_i(t, u_i, \boldsymbol{\theta}); \boldsymbol{\theta})]} Pr(u_i | T_i^* > t, \mathcal{Y}_i(t); \boldsymbol{\theta}) du_i, \end{aligned} \tag{18}$$

where $S(\cdot)$ is given in (13) and $\mathcal{M}_i(t) = \boldsymbol{H}_{it}^\top \boldsymbol{\delta} + \boldsymbol{Z}_{it}^\top \boldsymbol{u}_i$ in Equation (11).

According to Rizopoulos (2011), making direct estimation of Equation (18) is a difficult task by using the MLE of $\boldsymbol{\theta}$ and empirical Bayes estimate of u_i and the standard error or the estimate is also hard to derive. To avoid those problem, in stead of estimating $p_i(m|t)$ directly, we can calculate the posterior expectation of it by using MCMC technique and the posterior samples from the estimating algorithm that we developed in Section 3.2.2. Specifically, we are going to estimate

$$\begin{aligned} E_{\boldsymbol{\theta}|\mathcal{D}_n}[p_i(m|t)] &= Pr(T_i^* \geq m | T_i^* > t, \mathcal{Y}_i(t), \mathcal{D}_n) \\ &= \int Pr(T_i^* \geq m | T_i^* > t, \mathcal{Y}_i(t); \boldsymbol{\theta}) p(\boldsymbol{\theta} | \mathcal{D}_n) d\boldsymbol{\theta}. \end{aligned} \tag{19}$$

where the first part of the equation is given in (18).

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

Algorithm 1 MC algorithm to draw samples of $p_i(m|t)$

for k in $1 : K$ **do**

draw $\boldsymbol{\theta}^{(k)} \sim f(\boldsymbol{\theta}|\mathcal{D}_n)$

draw $u_i^{(k)} \sim f(u_i | T_i^* > t, \mathcal{Y}_i(t), \boldsymbol{\theta}^{(k)})$

compute $p_i^{(k)}(m|t) = S_i[m | \mathcal{M}_i(m, u_i^{(k)}, \boldsymbol{\theta}^{(k)}); \boldsymbol{\theta}^{(k)}] S_i[t | \mathcal{M}_i(t, u_i^{(k)}, \boldsymbol{\theta}^{(k)}); \boldsymbol{\theta}^{(k)}]^{-1}$

where K is the total number of MC iterations, $f(\boldsymbol{\theta}|\mathcal{D}_n)$ is the posterior distribution of $\boldsymbol{\theta}$ given in (15), and $f(u_i | T_i^*, \mathcal{Y}_i(t), \boldsymbol{\theta}^{(k)})$ is the posterior distribution of random effect for subject i if this subject is from the reference population. However, if the subject is a new patient that doesn't belong to our study population,

we need to do additional sampling for the random effect u_i from its posterior distribution using the posterior samples of the fixed effects collected from Gibbs sampler and its historical data. Upon collecting all of the K samples, the estimate of $p_i(m|t)$ can be calculated as the sample mean:

$$\hat{p}_i(m|t) = \frac{1}{K}\sum_{k=1}^{K} p_i^{(l)}(m|t), \tag{20}$$

or median and the standard error can be computed using the sample variance.

3.3.2 Validation of the Predictive Ability of the Longitudinal Biomarker

As discussed previously in Section 1.1.3, there are several methods developed to evaluate the accuracy of the predictions or how well the model can perform in predictions. In this thesis work, we will adopt the one from Rizopoulos (2011), namely the ROC based approach. This approach is designed to check the predictive ability of the longitudinal outcome in terms of discrimination between patients who will have the event from those who will not within a time interval of length Δt following the last longitudinal measurement taken at time t (Pencina et al., 2008). This is of medical relevance in practice, as it would provide a patient specific information to the physician as a reference to conduct personalized medical care.

In order to apply this method, we need to define the sensitivity and the specificity of the predictions. Following Rizopoulos (2011), the sensitivity is defined as

$$Pr[\mathcal{S}_i(t,k,\boldsymbol{c})|T_i^* > t, T_i^* \in (t, t+\Delta t]; \boldsymbol{\theta}], \tag{21}$$

and the specificity as

$$Pr[\mathcal{F}_i(t,k,\boldsymbol{c})|T_i^* > t, T_i^* > t+\Delta t; \boldsymbol{\theta}]. \tag{22}$$

In those definitions, $\mathcal{S}_i(t,k,\boldsymbol{c}) = \{Y_i(s) \leq c_s, k \leq s \leq t\}$ is defined as success (or event) and $\mathcal{F}_i(t,k,\boldsymbol{c}) = \mathbb{R}^{n(k,t)} \setminus \{Y_i(s) \leq c_s, k \leq s \leq t\}$ is defined as failure, where $\boldsymbol{c}$ is a vector of threshold values and c_s is the threshold value at time s, $\mathbb{R}^r$ denotes the $r-$dimensional Euclidean space and $n(k,t)$ is the total number of longitudinal measurements in interval $[k,t]$. Note that in above definitions, the default choice is that smaller longitudinal measurement leads to higher risk, however, this setting can be reversed in practice whenever it

is necessary.

To obtain the estimation of (21) and (22), we will use a similar MC technique as Algorithm 1. Take the estimation approach for sensitivity as an example. First of all, Equation (21) can be expressed as (by omitting the parameters and covariates)

$$Pr\{\mathcal{S}_i(t,k,\boldsymbol{c})|T_i^*>t, T_i^*\in(t,t+\Delta t]\}=\frac{Pr\{\mathcal{S}_i(t,k,\boldsymbol{c}), T_i^*\in(t,t+\Delta t]|T_i^*>t\}}{1-Pr(T_i^*>t+\Delta t|T_i^*>t)}$$

The numerator can be rephrased as

$$\begin{aligned} Pr\{\mathcal{S}_i(t,k,\boldsymbol{c}), T_i^*\in(t,t+\Delta t]|T_i^*>t\} &= \int Pr\{\mathcal{S}_i(t,k,\boldsymbol{c}), T_i^*\in(t,t+\Delta t]|T_i^*>t,u_i\} \\ &\times \quad p(u_i|T_i^*>t)du_i \\ &= \int\left\{\prod_{s=k}^{t}\Phi\left[\frac{c_s-\omega_i(s)}{\xi}\right]\right\} \\ &\times \left[1-\frac{\mathcal{S}_i\{t+\Delta t|\mathcal{M}_i(t+\Delta t,u_i)\}}{\mathcal{S}_i\{t|\mathcal{M}_i(t,u_i)\}}\right]\times p(u_i|T_i^*>t)du_i \quad (23) \end{aligned}$$

where $\Phi(\cdot)$ is the cumulative distribution function of standard normal density, $\omega_i(s)$ and ξ are the mean and standard deviation respectively in the location-scale representation of $Y_i(s)$ in (11), assuming ALD(0, σ,τ) of the error term.

Similarly, the denominator can be written as

$$1-Pr(T_i^*>t+\Delta t|T_i^*>t)=1-\int\frac{\mathcal{S}_i\{t+\Delta t|\mathcal{M}_i(t+\Delta t,u_i)\}}{\mathcal{S}_i\{t|\mathcal{M}_i(t,u_i)\}}p(u_i|T_i^*>t)du_i. \quad (24)$$

For simplicity in the notation, let $\mathcal{E}_1(u_i,\boldsymbol{\theta})=\left\{\prod_{s=k}^{t}\Phi\left[\frac{c_s-\omega_i(s)}{\xi}\right]\right\}\left[1-\frac{\mathcal{S}_i\{t+\Delta t|\mathcal{M}_i(t+\Delta t,u_i)\}}{\mathcal{S}_i\{t|\mathcal{M}_i(t,u_i)\}}\right]$ and $\mathcal{E}_2(u_i,\boldsymbol{\theta})=\mathcal{S}_i\{t+\Delta t|\mathcal{M}_i(t+\Delta t,u_i)\}\mathcal{S}_i\{t|\mathcal{M}_i(t,u_i)\}^{-1}$, then the sensitivity can be written in terms of the expectations of $\mathcal{E}_1(u_i,\boldsymbol{\theta})$ and $\mathcal{E}_2(u_i,\boldsymbol{\theta})$ with respect to the marginal posterior distribution of the random effects, i.e. $p(u_i|T_i^*>t)$. Note that

$$p(u_i|T_i^*>t)\propto\int p(\mathcal{Y}_i(t)|u_i)\mathcal{S}_i\{t|\mathcal{M}_i(t,b_i)\}p(u_i)d\mathcal{Y}_i(t) \quad (25)$$

Based on above derivations, we now can develop the following MC algorithm to simulate samples of the sensitivity:

Algorithm 2 MC algorithm to compute sensitivity of the predictions

for k in $1:K$ **do**

draw $\boldsymbol{\theta}^{(k)} \sim f(\boldsymbol{\theta}|\mathcal{D}_n)$

draw $\mathcal{Y}_i^{(k)}(t) \sim N(\boldsymbol{X}_{it}^\top \boldsymbol{\beta}^{(k)} + \boldsymbol{H}_{it}^\top \boldsymbol{\delta}^{(k)} + \boldsymbol{Z}_{it}^\top \boldsymbol{u}_i^{(k-1)} + \kappa_1 e_{ij}, \kappa_2^2 \sigma^{(k)} e_{ij})$

draw $u_i^{(k)} \sim f(u_i|T_i^* > t, \mathcal{Y}_i^{(k)}(t), \boldsymbol{\theta}^{(k)})$

compute $\mathcal{E}_1(u_i^{(k)}, \boldsymbol{\theta}^{(k)})$ and $\mathcal{E}_2(u_i^{(k)}, \boldsymbol{\theta}^{(k)})$

Differently as in Algorithm 1, here the random effects have to be redrawn no matter the subject belongs to study population or not as it is based on the new longitudinal value that is drawn in step 2 in Algorithm 2. Once we get K realizations of $\mathcal{E}_1(u_i, \boldsymbol{\theta})$ and $\mathcal{E}_2(u_i, \boldsymbol{\theta})$, we are ready to calculate the MC estimate of the sensitivity as follows:

$$\widehat{Pr}[\mathcal{S}_i(t,k,\boldsymbol{c})|T_i^* > t, T_i^* \in (t, t+\Delta t]; \boldsymbol{\theta}] = \frac{\sum_{k=1}^K \mathcal{E}_1(u_i^{(k)}, \boldsymbol{\theta}^{(k)})/K}{1 - \sum_{k=1}^K \mathcal{E}_2(u_i^{(k)}, \boldsymbol{\theta}^{(k)})/K}. \tag{26}$$

The standard error of the estimate can also be obtained from the values calcuated based on the sample samples subsequently.

Specificity can be estimated in a similar manner. Once the estimates of sensitivity and specificity are available, we can construct the ROC curve and calculate the area under the curve (AUC) for some specific time interval Δt over the follow-up period as the overall evaluation of the capacity of the predictive model proposed.

4 Plan for Simulation Studies

Simulation studies are designed mainly for two aims. First, to check the validity of our proposed fully Bayesian algorithm in estimating the regression coefficients in the new JM framework. Our focus of the simulation results lies on the accuracy (i.e. bias) and the precision (i.e. standard deviation) of our estimations based on the posterior distributions. The second simulation study is planned to check the performance of our proposed estimator of the survival probability given in (20). This can be done by comparing our estimates with the "gold standard", which is calculated based on the true (simulated) value of random effects, fixed

effects as well as the parameters.

4.1 Simulation Study 1 (Estimation Accuracy)

The first simulation study is designed to to check the accuracy of our proposed fully Bayesian estimating algorithm, which is the crucial basis for future prediction performance. Following Farcomeni and Viviani (2014), by varying the values of the association parameters α_1 and α_2 in our model (11), we will have four different settings for simulation study, which are:

1. $(\alpha_1, \alpha_2) = (0, 0)$, the two models are indepdent with each other
2. $(\alpha_1, \alpha_2) = (1, 0)$, the two models are related only through the observed heterogeneity in some covarites, i.e. $\boldsymbol{H}_{it}$ in our model
3. $(\alpha_1, \alpha_2) = (0, 1)$, the two models are related only through the unobserved heterogeneity, i.e. the random effects
4. $(\alpha_1, \alpha_2) = (1, 1)$, the depdence of the two models is explained by both observed and unobserved heterogeneity

Under different combinations of α_1 and α_2 values, we choose the regression coefficients $\boldsymbol{\beta}, \boldsymbol{\delta}, \boldsymbol{\gamma} = (1, 1)^\top$, the covriates $\boldsymbol{Z}_{it} = (1, t)^\top$, $\boldsymbol{H}_{it} = (h_{i1}, h_{i2} * t)^\top$, $\boldsymbol{X}_{it} = (1, x_i)^\top$, and $\boldsymbol{W}_i = (w_{i1}, w_{i2})^\top$ with $h_{i1}, h_{i2}, x_i, w_{i1}$ and w_{i2} generated from independent standard normal distributions, and the random effects $\boldsymbol{u}_i$ from bivariate normal with mean 0, standard deviations equal 0.3 and correlation 0.16. We also fix $\sigma = 1$ and vary the quantile τ among $\{0.25, 0.5, 0.75\}$ for the ALD specification when simulating longitudinal data.

To simulate the survival time data, for simplicity, we fix $h_0(s) = 1$ and obtain the survival distribution as

$$S(t|\boldsymbol{u}_i, \boldsymbol{H}_{it}, \boldsymbol{W}_i) = \exp\left\{-\frac{e^{\alpha_1(\delta_1 H_{i1} + \delta_2 H_{i2t}) + \alpha_2(u_{i1} + u_{i2}t) + \boldsymbol{\gamma}^\top \boldsymbol{W}_i} - e^{\alpha_1 \delta_1 H_{i1} + \alpha_2 u_{i1} + \boldsymbol{\gamma}^\top \boldsymbol{W}_i}}{\alpha_2 u_{i2} + \alpha_1 \delta_2 h_{i2}}\right\}$$

when $\alpha_1 \neq 0$ or $\alpha_2 \neq 0$ and

$$S(t|\boldsymbol{u}_i, \boldsymbol{H}_{it}, \boldsymbol{W}_i) = \exp\{-te^{\boldsymbol{\gamma}^\top \boldsymbol{W}_i}\}$$

when $\alpha_1 = \alpha_2 = 0$. We then can obtain event time T_i by inverting above survival function after generating n random variates from standard uniform distribution. To obtain a censoring proportion around 25%, we

choose the censoring time $C_i/5$ be distributed according to $beta(4,1)$.

To simulate the longitudinal data, we draw them independently from the ALD for the τth quantile, centered on

$$\boldsymbol{\beta}^\top \boldsymbol{X}_{it} + \boldsymbol{\delta}^\top \boldsymbol{H}_{it} + \boldsymbol{u}_i^\top \boldsymbol{Z}_{it},$$

and with dispersion parameter σ. We keep maximum six observations for each subject at follow-up time $t = (0, 0.25, 0.5, 0.75, 1, 3)$ respectively, after incorporating the drop-out information. Results of Simulation 1 will be put in Table 1.

4.2 Simulation Study 2 (Predictive Performance)

In this simulation study, we will use the true simulated random effects, covariates and coefficients to calculate the survival probability for subject i as the "gold standard", which is given by

$$\frac{S_i[m|\mathcal{M}_i(m, u_i, \boldsymbol{\theta}); \boldsymbol{\theta}]}{S_i[t|\mathcal{M}_i(t, u_i, \boldsymbol{\theta}); \boldsymbol{\theta}]}.$$

We then use Algorithm 1 to computed the predicted survival probability for the same subject. By doing this for many different selected subjects, we will be able to evaluate how "close" the predicted values from our proposed model are to the true values by using Bland-Altman plot (Bland and Altman, 1986), which is a plotting method used to assess the agreement of the results between two measurement methods. We will do this simulation study by varying the follow-up time, i.e. t, in longitudinal process as well as the prediction time interval, i.e. Δt, in the model specification to mimic what is happening in the real-world studies.

Table 1: Bias and standard error of the parameter estimates from proposed fully Bayesian estimating method

			τ=0.25													
			β_1		β_2		δ_1		δ_2		γ_1		γ_2		α_1		α_2	
n	α_1	α_2	bias	s.d.	bias	s.d.	bias	s.d.	bias	s.d.	bias	s.d.	bias	s.d.	bias	s.d.	bias	s.d.
500	0	0																
500	1	0																
500	0	1																
500	1	1																
			τ=0.5													
			β_1		β_2		δ_1		δ_2		γ_1		γ_2		α_1		α_2	
n	α_1	α_2	bias	s.d.	bias	s.d.	bias	s.d.	bias	s.d.	bias	s.d.	bias	s.d.	bias	s.d.	bias	s.d.
500	0	0																
500	1	0																
500	0	1																
500	1	1																
			τ=0.75													
			β_1		β_2		δ_1		δ_2		γ_1		γ_2		α_1		α_2	
n	α_1	α_2	bias	s.d.	bias	s.d.	bias	s.d.	bias	s.d.	bias	s.d.	bias	s.d.	bias	s.d.	bias	s.d.
500	0	0																
500	1	0																
500	0	1																
500	1	1																

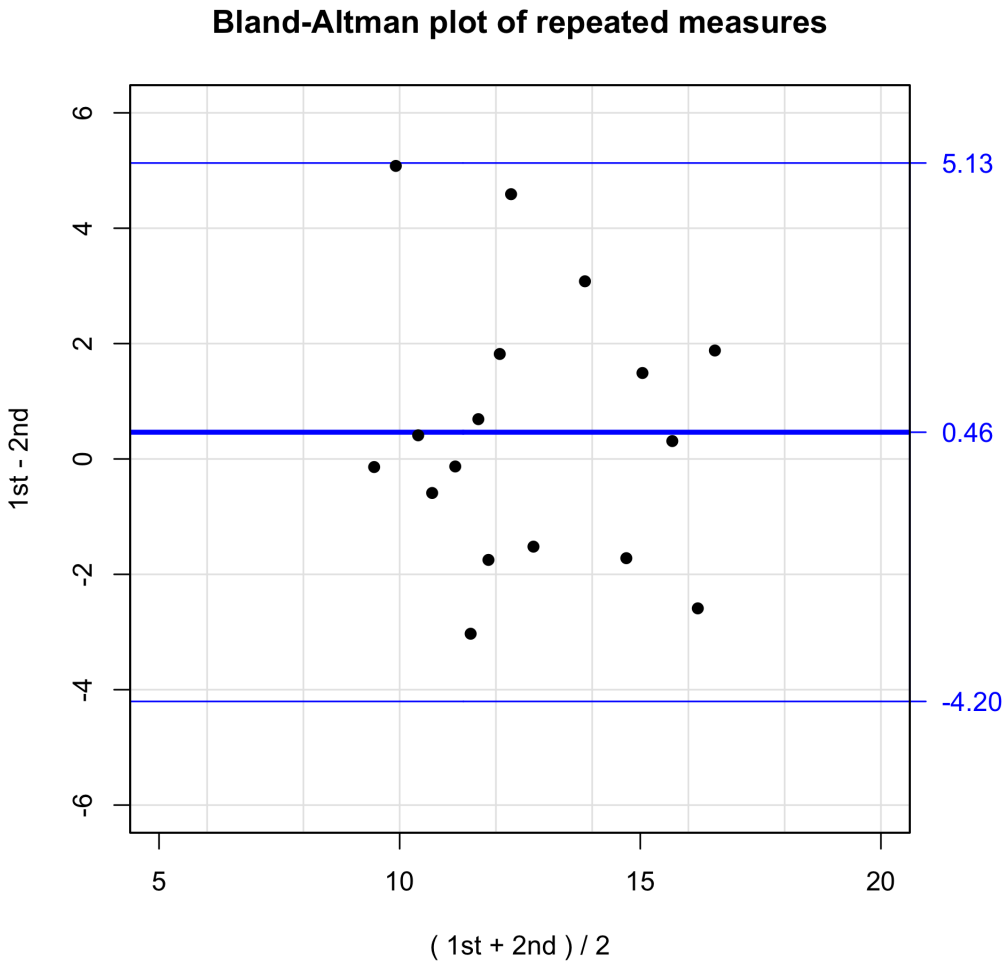

Figure 2: An example of Bland-Altman plot

Table 2: Summary table of comparing the predictive results from proposed method with gold standard

	$\Delta t = 2$	$\Delta t = 4$	$\Delta t = 6$
	bias(lower, upper)	bias(lower, upper)	bias(lower, upper)
t=2			
t=4			
t=8			
t=16			

References

J Martin Bland and Douglas G Altman. Statistical methods for assessing agreement between two methods of clinical measurement. *The lancet*, 327(8476):307–310, 1986.

Elizabeth R Brown, Joseph G Ibrahim, and Victor DeGruttola. A flexible b-spline model for multiple longitudinal biomarkers and survival. *Biometrics*, 61(1):64–73, 2005.

William C Cushman, Barry R Davis, Sara L Pressel, Jeffrey A Cutler, Paula T Einhorn, Charles E Ford, Suzanne Oparil, Jeffrey L Probstfield, Paul K Whelton, Jackson T Wright, et al. Mortality and morbidity during and after the antihypertensive and lipid-lowering treatment to prevent heart attack trial. *The Journal of Clinical Hypertension*, 14(1):20–31, 2012.

Barry R Davis, Jeffrey A Cutler, David J Gordon, Curt D Furberg, Jackson T Wright, William C Cushman, Richard H Grimm, John LaRosa, Paul K Whelton, H Mitchell Perry, et al. Rationale and design for the antihypertensive and lipid lowering treatment to prevent heart attack trial (allhat). *American Journal of Hypertension*, 9(4):342–360, 1996.

Peter Diggle, Patrick Heagerty, Kung-Yee Liang, and Scott Zeger. *Analysis of longitudinal data*. Oxford University Press, 2002.

Robert M Elashoff, Gang Li, and Ning Li. A joint model for longitudinal measurements and survival data in the presence of multiple failure types. *Biometrics*, 64(3):762–771, 2008.

Alessio Farcomeni and Sara Viviani. Longitudinal quantile regression in the presence of informative dropout through longitudinal–survival joint modeling. *Statistics in medicine*, 2014.

Cheryl L Faucett and Duncan C Thomas. Simultaneously modelling censored survival data and repeatedly measured covariates: a gibbs sampling approach. *Statistics in medicine*, 15(15):1663–1685, 1996.

US FDA. Paving the way for personalized medicine. fda's role in a new era of medical product development., 2013.

Liya Fu and You-Gan Wang. Quantile regression for longitudinal data with a working correlation model. *Computational Statistics & Data Analysis*, 56(8):2526–2538, 2012.

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

Richard H Grimm, Karen L Margolis, Vasilios Papademetriou, William C Cushman, Charles E Ford, Judy Bettencourt, Michael H Alderman, Jan N Basile, Henry R Black, Vincent DeQuattro, et al. Baseline characteristics of participants in the antihypertensive and lipid lowering treatment to prevent heart attack trial (allhat). *Hypertension*, 37(1):19–27, 2001.

Xu Guo and Bradley P Carlin. Separate and joint modeling of longitudinal and event time data using standard computer packages. *The American Statistician*, 58(1):16–24, 2004.

Robin Henderson, Peter Diggle, and Angela Dobson. Joint modelling of longitudinal measurements and event time data. *Biostatistics*, 1(4):465–480, 2000.

Sin-Ho Jung. Quasi-likelihood for median regression models. *Journal of the American Statistical Association*, 91(433):251–257, 1996.

Roger Koenker. *Quantile regression*. Cambridge university press, 2005.

Roger Koenker and Gilbert Bassett Jr. Regression quantiles. *Econometrica: journal of the Econometric Society*, pages 33–50, 1978.

Roger Koenker and Kevin Hallock. Quantile regression: An introduction. *Journal of Economic Perspectives*, 15(4):43–56, 2001.

Roger Koenker and Jose AF Machado. Goodness of fit and related inference processes for quantile regression. *Journal of the american statistical association*, 94(448):1296–1310, 1999.

Samuel Kotz, Tomasz Kozubowski, and Krzysztof Podgorski. *The Laplace Distribution and Generalizations: A Revisit With Applications to Communications, Exonomics, Engineering, and Finance*. Number 183. Springer, 2001.

Hideo Kozumi and Genya Kobayashi. Gibbs sampling methods for bayesian quantile regression. *Journal of statistical computation and simulation*, 81(11):1565–1578, 2011.

Nan M Laird and James H Ware. Random-effects models for longitudinal data. *Biometrics*, pages 963–974, 1982.

Haiqun Lin, Bruce W Turnbull, Charles E McCulloch, and Elizabeth H Slate. Latent class models for joint analysis of longitudinal biomarker and event process data: application to longitudinal prostate-specific antigen readings and prostate cancer. *Journal of the American Statistical Association*, 97(457):53–65, 2002.

Yuan Liu and Matteo Bottai. Mixed-effects models for conditional quantiles with longitudinal data. *The International Journal of Biostatistics*, 5(1), 2009.

Youxi Luo, Heng Lian, and Maozai Tian. Bayesian quantile regression for longitudinal data models. *Journal of Statistical Computation and Simulation*, 82(11):1635–1649, 2012.

Michael J Pencina, Ralph B D'Agostino, and Ramachandran S Vasan. Evaluating the added predictive ability of a new marker: from area under the roc curve to reclassification and beyond. *Statistics in medicine*, 27 (2):157–172, 2008.

Linda B Piller, Sarah Baraniuk, Lara M Simpson, William C Cushman, Barry M Massie, Paula T Einhorn, Suzanne Oparil, Charles E Ford, James F Graumlich, Richard A Dart, et al. Long-term follow-up of participants with heart failure in the antihypertensive and lipid-lowering treatment to prevent heart attack trial (allhat). *Circulation*, 124(17):1811–1818, 2011.

Martyn Plummer et al. Jags: A program for analysis of bayesian graphical models using gibbs sampling. In *Proceedings of the 3rd International Workshop on Distributed Statistical Computing (DSC 2003). March*, pages 20–22, 2003.

Cécile Proust-Lima and Jeremy MG Taylor. Development and validation of a dynamic prognostic tool for prostate cancer recurrence using repeated measures of posttreatment psa: a joint modeling approach. *Biostatistics*, 10(3):535–549, 2009.

Dimitris Rizopoulos. Dynamic predictions and prospective accuracy in joint models for longitudinal and time-to-event data. *Biometrics*, 67(3):819–829, 2011.

Dimitris Rizopoulos and Pulak Ghosh. A bayesian semiparametric multivariate joint model for multiple longitudinal outcomes and a time-to-event. *Statistics in medicine*, 30(12):1366–1380, 2011.

Jeremy MG Taylor, Yongseok Park, Donna P Ankerst, Cecile Proust-Lima, Scott Williams, Larry Kestin, Kyoungwha Bae, Tom Pickles, and Howard Sandler. Real-time individual predictions of prostate cancer recurrence using joint models. *Biometrics*, 69(1):206–213, 2013.

AA Tsiatis, Victor Degruttola, and MS Wulfsohn. Modeling the relationship of survival to longitudinal data measured with error. applications to survival and cd4 counts in patients with aids. *Journal of the American Statistical Association*, 90(429):27–37, 1995.

Yan Wang and Jeremy M G Taylor. Jointly modeling longitudinal and event time data with application to acquired immunodeficiency syndrome. *Journal of the American Statistical Association*, 96(455):895–905, 2001.

Jane Xu and Scott L Zeger. Joint analysis of longitudinal data comprising repeated measures and times to events. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 50(3):375–387, 2001.

Keming Yu and Rana A Moyeed. Bayesian quantile regression. *Statistics & Probability Letters*, 54(4): 437–447, 2001.

Menggang Yu, Ngayee J Law, Jeremy MG Taylor, and Howard M Sandler. Joint longitudinal-survival-cure models and their application to prostate cancer. *Statistica Sinica*, 14(3):835–862, 2004.

Menggang Yu, Jeremy M G Taylor, and Howard M Sandler. Individual prediction in prostate cancer studies using a joint longitudinal survival–cure model. *Journal of the American Statistical Association*, 103(481): 178–187, 2008.