1 Statistical Methods

In this section we introduce three major statistical methods that will be used in the projects, including Bayesian quantile regression, joint model uses longitudinal quantile regression and dynamic prediction method for the probability of future events based on joint model of longitudinal and survival data.

1.1 Bayesian Quantile Regression

1.1.1 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 τ^{th} 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 τ^{th} quantile can be obtained by solving

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

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 solve (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).

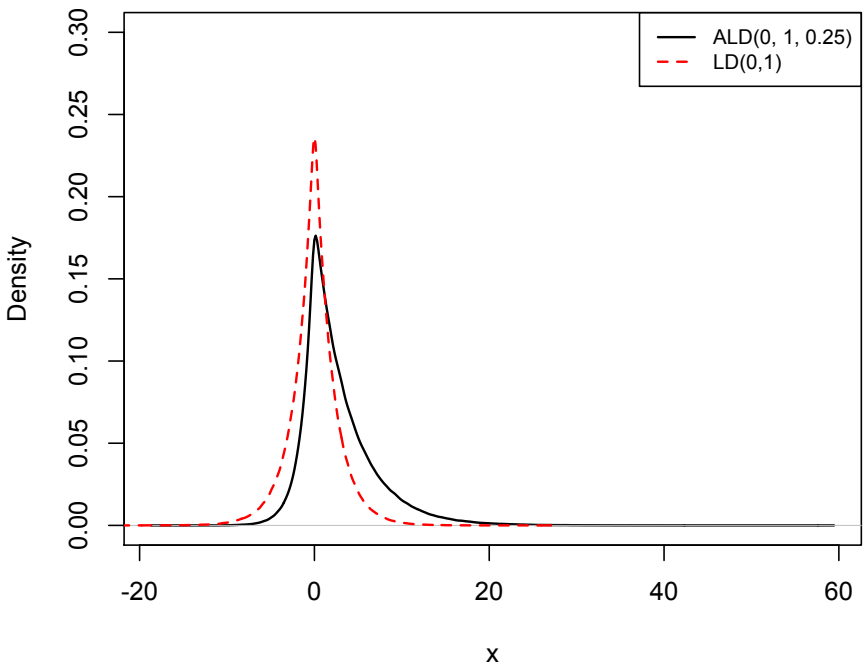

Figure 1: Graphical visualization of ALD versus Laplace distribution (LD).

1.1.2 Bayesian Linear Quantile Mixed Model

As a natural extension to linear quantile regression, 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. Under the assumption that the random error follows ALD$(0, \sigma, \tau)$ distribution, 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 this assumption the random error is represented as $\varepsilon_{ij} = \varrho_1 e_{ij} + \varrho_2\sqrt{\sigma e_{ij}}v_{ij}$,

where

$$\varrho_1 = \frac{1-2\tau}{\tau(1-\tau)}, \text{ and } \varrho_2^2 = \frac{2}{\tau(1-\tau)},$$

and

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

The linear mixed model is then reparameterized as

$$Y_{ij} = \boldsymbol{X}_{ij}^\top \boldsymbol{\beta} + \boldsymbol{Z}_{ij}^\top \boldsymbol{u}_i + \varrho_1 e_{ij} + \varrho_2 \sqrt{\sigma e_{ij}} v_{ij}, \tag{9}$$

or equivalently

$$f(Y_{ij}|\boldsymbol{\beta}, \boldsymbol{u}_i, e_{ij}, \sigma) = \frac{1}{\sqrt{2\pi\varrho_2^2\sigma e_{ij}}} \exp\left[-\frac{1}{2\varrho_2^2\sigma e_{ij}}(Y_{ij} - \boldsymbol{X}_{ij}^\top\boldsymbol{\beta} - \boldsymbol{Z}_{ij}^\top\boldsymbol{u}_i - \varrho_1 e_{ij})^2\right]. \tag{10}$$

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

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

1.2 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, we 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 i^{th} subject 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, the degree of associations from those two sources of measurements (observed and unobserved) are measure by another two parameters α_1 and α_2, respectively.

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

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

where the first equation is the linear quantile mixed model discussed in Section 1.1.2 and the second equation takes the format of Cox proportional hazards model where $h_0(\cdot)$ is the baseline hazard function.

In the model, individual heterogeneity is captured by the term $\boldsymbol{Z}_{it}^{\top}\boldsymbol{u}_i$, which is the deviation of subject i from the population average. The posterior estimates of the random effects can be used to draw subject specific prediction of future event probability or longitudinal outcome.

Also note that in quantile regression the parameter estimates are function of the quantiles. 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 iteratively, collect and compare the results for different quantiles. In contrast, if the interest only lies in investigating the effect of lower of higher quantile for the longitudinal outcome on the survival probability, we may just fix the quantile to a specific value and do the analysis.

1.2.1 The Survival Model

As the details for the longitudinal model has been discussed previously in Section (1.1.2), here we will only focus on the survival component of the JM.

For i^{th} subject, 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 $S(\cdot)$ is the survival function, i.e.

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

and $h(T_i|\mathcal{T}_{iT_i}, \boldsymbol{W}_i)$ is given in (11). For the baseline hazard $h_0(s)$, parametric form like Weibull model can be used. However, the choice of baseline hazard is not a main focus of the this project.

1.2.2 Complete Likelihood Function and Bayesian Inference

The complete joint likelihood for longitudinal and survival outcomes, assuming random effects are known, for the i^{th} 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 (13) 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 complex derivation of the expectation and maximization functions as well as obtaining the standard error of the estimates, we in stead 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:

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

Similarly as in Section (1.1.2), we choose the following priors for the parameters

$$\begin{gathered}
\boldsymbol{\beta} \sim MVN_p(\boldsymbol{b}_0, \boldsymbol{B_0}), \\
\boldsymbol{\delta} \sim MVN_k(\boldsymbol{d}_0, \boldsymbol{D_0}), \\
\boldsymbol{\gamma} \sim MVN_k(\boldsymbol{g}_0, \boldsymbol{G_0}), \\
\alpha_1, \alpha_2 \sim N(a_0, \sigma_a), \\
\sigma \sim IG(s_0, s_1).
\end{gathered}$$

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:

$$\begin{gathered}
w_{ii} \sim unif(a, b), \\
w_{ij} \sim N(\mu_w, \sigma_w), \text{ for } i \neq j.
\end{gathered}$$

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.

1.3 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 survival probabilities. The accuracy 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 $\boldsymbol{\theta}_\tau$.

1.3.1 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 new subject i 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 implicit 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 the reference population of size n, based on which we fit out 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 1.2.2. 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 in 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. Upon collecting all of the K samples, the estimates of $p_i(m|t)$ can be calculated as the sample mean:

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

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

1.3.2 Validation of the Prediction Accuracy

As discussed previously in Section **??**, there are several methods developed to evaluated the accuracy of the predictions. In our work, we will adopt the one from Rizopoulos (2011), namely the ROC based approach. This approach is designed to discriminate 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 it the threshold value at time s, $\mathbb{R}^r$ is $r-$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

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

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

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

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

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

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 $\mathcal{Y}_i^{(k)}(t) \sim N(\boldsymbol{X}_{it}^\top \boldsymbol{\beta}^{(k)} + \boldsymbol{H}_{it}^\top \boldsymbol{\delta}^{(k)} + \boldsymbol{Z}_{it}^\top \boldsymbol{u}_i^{(k-1)} + \varrho_1 e_{ij}, \varrho_2^2 \sigma^{(k)} e_{ij})$
- 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 $\mathcal{E}_2(u_i^{(k)}, \boldsymbol{\theta}^{(k)})$

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:

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

According to the multivariate version of Delta method, we can also obtain the standard error for the estimate, which is given by

$$s.e.(\widehat{Pr}[\mathcal{S}_i(t, k, \boldsymbol{c}) | T_i^* > t, T_i^* \in (t, t + \Delta t]; \boldsymbol{\theta}]) = (gVg^\top)^{1/2}, \tag{27}$$

where

$$g = \left(\frac{1}{1 - \sum_{k=1}^{K} \mathcal{E}_2(u_i^{(k)}, \boldsymbol{\theta}^{(k)})/K}, \frac{\sum_{k=1}^{K} \mathcal{E}_1(u_i^{(k)}, \boldsymbol{\theta}^{(k)})/K}{(1 - \sum_{k=1}^{K} \mathcal{E}_2(u_i^{(k)}, \boldsymbol{\theta}^{(k)})/K)^2}\right),$$

and

$$V = K^{-1} \begin{bmatrix} var(\mathcal{E}_1(u_i^{(k)}) & cov(\mathcal{E}_1(u_i^{(k)}, \mathcal{E}_2(u_i^{(k)}) \\ cov(\mathcal{E}_1(u_i^{(k)}, \mathcal{E}_2(u_i^{(k)}) & var(\mathcal{E}_2(u_i^{(k)}) \end{bmatrix}.$$

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.

References

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

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

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.

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.

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.

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

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