Dynamic Disease Screening by Joint Modeling of Survival and Longitudinal Data

Peihua Qiu[1] and Lu You[2]

[1]Department of Biostatistics, University of Florida, Gainesville, FL 32610

[2]Health Informatics Institute, University of South Florida, Tampa, FL 33612

Abstract

Sequential monitoring of dynamic processes is an active research area because of its broad applications in different industries and scientific research projects, including disease screening in medical research. In the literature, it has been shown that dynamic screening system (DySS) is a powerful tool for sequential monitoring of dynamic processes. To detect a disease (e.g., stroke) for a patient, existing DySS methods first estimate the regular longitudinal pattern of certain disease predictors (e.g., blood pressure, cholesterol level) from an in-control (IC) dataset that contains observations of a group of non-diseased people, and then compare the longitudinal pattern of the observed disease predictors of the given patient with the estimated regular longitudinal pattern. A signal of disease occurrence is triggered if their cumulative difference exceeds a certain level, facilitated by a built-in control chart. In practice, a dataset containing longitudinal observations of the disease predictors of both non-diseased and diseased people is often available in advance, from which it is possible to explore the relationship between the disease occurrence and the longitudinal pattern of the disease predictors. This relationship should be helpful for disease screening. In this paper, a new DySS method is suggested based on this idea. Numerical studies confirm that it can improve the existing DySS methods for disease screening.

Key Words: Disease predictors; Dynamic processes; Joint modeling; Longitudinal data; Survival analysis; Time to event.

1 Introduction

Disease early detection is critically important for our wellbeing. To detect a disease (e.g., stroke) in a timely manner, we need to have clinical visits and check certain disease predictors (e.g., blood pressure, cholesterol level) in a regular basis. The repeated measures of the disease predictors can be regarded as sequential observations of an underlying process, and this process is dynamic in the

sense that its distribution would change over time even when the process is in control (IC) (i.e., the related person is healthy and does not have the disease in question). This paper aims to develop a new methodology to detect a disease early by sequentially monitoring the dynamic process of the disease predictors.

A large amount of existing literature has been devoted to modeling longitudinal data, among which nonparametric approaches have received increasing attention since parametric models with stringent assumptions may not be able to describe the observed data well. Examples of nonparametric models for analyzing longitudinal data include those discussed in Shi et al. (1996), Brumback and Rice (1998), Wu and Zhang (2002) and Xiang et al. (2013). There is also an increasing literature on identifying disease predictors that are associated with the survival outcomes related to a disease (e.g., Wulfsohn and Tsiatis, 1997; Brown and Ibrahim, 2003; Chi and Ibrahim, 2006; Rizopoulos and Ghosh, 2011). Nonetheless, all these methods are retrospective, and thus cannot handle the prospective disease early detection problem effectively. To overcome this limitation, Qiu and Xiang (2014, 2015) suggested the dynamic screening system (DySS) approach for disease early detection, which combined the strengths of longitudinal data modeling approaches and sequential process monitoring approaches. By a DySS method, the regular longitudinal pattern of the disease predictors is first estimated by a longitudinal data modeling approach from an IC dataset that contains observations of a group of non-diseased people. Then, to detect the disease for a given person, his/her longitudinal pattern of the observed disease predictors is compared with the estimated regular longitudinal pattern, and a signal of disease occurrence is triggered if their cumulative difference exceeds a certain level, facilitated by a built-in statistical process control (SPC) chart. Some modified versions of this approach can be found in Li and Qiu (2016, 2017), Qiu et al. (2018) and You and Qiu (2019). However, these methods only use the information in the observed longitudinal data of the disease predictors for disease early detection, and the association between the longitudinal data and the survival outcomes related to the occurrence of the disease in question is ignored completely. For this reason, You and Qiu (2020) recently proposed a new disease early detection method that attempted to quantify the relationship between the longitudinal data and the survival outcomes by a proportional hazards model, from a training data containing longitudinal observations of the disease predictors of both non-diseased and diseased people. In the literature, it has been well demonstrated that joint modeling of the longitudinal and survival data is often more effective than the corresponding survival model alone that includes the related

longitudinal covariates (cf., Wulfsohn and Tsiatis, 1997). Based on this intuition, we develop a new disease early detection method in this paper using the joint modeling approach. Numerical studies show that the new method is indeed more effective than some representative existing methods.

The remainder of the paper is organized as follows. Joint modeling of multivariate longitudinal data and survival data is discussed in Section 2. A new disease early detection method based on the joint modeling approach is described in Section 3. Several extensions of the proposed method are discussed in Section 4. Some numerical results regarding its performance are presented in Section 5. A real-data example for demonstrating its application is discussed in Section 6. Several remarks conclude the paper in Section 7.

2 A Joint Model and Its Estimation

A joint model. Suppose there is a training dataset containing longitudinal observations of p disease predictors and survival outcomes related to the occurrence of a given disease of interest of m individuals. For the ith individual, longitudinal observations of the kth disease predictor are obtained at times $t_{ik1}, t_{ik2}, \ldots, t_{ikn_{ik}}$, for $i = 1, \ldots, m$ and $k = 1, \ldots, p$, where the observation times of different disease predictors could be different. The corresponding observations of the kth disease predictor are denoted as $y_{ik}(t_{ik1}), y_{ik}(t_{ik2}), \ldots, y_{ik}(t_{ikn_{ik}})$. The survival outcome of the ith individual is denoted by (T_i, Δ_i), where T_i represents the last follow-up time and Δ_i is the indicator of the event that the ith individual has the disease at T_i (i.e., $\Delta_i = 1$ denotes the event being observed, and $\Delta_i = 0$ otherwise). As in the classical survival model setup, T_i is assumed to be the minimum of the event time D_i and the censoring time C_i. Then, $\Delta_i = I(D_i \leq C_i)$. Let $N_i(t) = \Delta_i I(T_i \leq t)$ be the right-continuous counting process of the survival status for the ith individual, $R(t) = \{i : T_i \geq t\}$ be the set of individuals who are still at risk at time t, and $[0, \mathcal{T}]$ be the study period that contains all observation times. Then, the observed longitudinal observations of the disease predictors are assumed to follow the multivariate nonparametric mixed-effects model below: for $j = 1, \ldots, n_{ik}$, $i = 1, \ldots, m$ and $k = 1, \ldots, p$,

$$y_{ik}(t_{ikj}) = \mu_k^{(0)}(t_{ikj}) + \Delta_i \delta_k(t_{ikj}) + v_{ik}(t_{ikj}) + \epsilon_{ik}(t_{ikj}), \tag{1}$$

where $\mu_k^{(0)}(t) = \mathrm{E}[y_{ik}(t) | t \leq T_i, \Delta_i = 0]$ is the population mean function of the kth disease predictor for individuals without the disease in concern, $\mu_k^{(1)}(t) = \mu_k^{(0)}(t) + \delta_k(t) = \mathrm{E}[y_{ik}(t) | t \leq T_i, \Delta_i =$

1] is the population mean function of the kth disease predictor for individuals with the disease observed during the study period, $v_{ik}(t)$ is the zero-mean random-effects function for describing the person-to-person variation from the population mean $\mu_k^{(0)}(t) + \Delta_i \delta_k(t)$, and $\epsilon_{ik}(t)$ is the i.i.d. pure measurement error with $\mathrm{Var}(\epsilon_{ik}(t)) = \sigma_k^2(t)$, for each i. Let $\mathbf{v}_i(t) = (v_{i1}(t), \ldots, v_{ip}(t))^\mathsf{T}$, $\boldsymbol{\epsilon}_i(t) = (\epsilon_{i1}(t), \ldots, \epsilon_{ip}(t))^\mathsf{T}$, $\boldsymbol{\mu}^{(0)}(t) = (\mu_1^{(0)}(t), \ldots, \mu_p^{(0)}(t))^\mathsf{T}$, and $\boldsymbol{\delta}(t) = (\delta_1(t), \ldots, \delta_k(t))^\mathsf{T}$. Then, $\mathbf{m}_i(t) = (m_{i1}(t), \ldots, m_{ip}(t))^\mathsf{T} = \boldsymbol{\mu}^{(0)}(t) + \Delta_i \boldsymbol{\delta}(t) + \mathbf{v}_i(t)$ denotes the latent trajectories of the disease predictors after the pure measurement errors are removed from their longitudinal observations. To describe the association between these latent trajectories and the observed survival outcomes, the following proportional hazards model is considered:

$$\lambda_i(t) = \lambda_0(t) \exp\{\boldsymbol{\beta}^\mathsf{T} \mathbf{m}_i(t)\}, \qquad \text{for } i = 1, \ldots, m, \tag{2}$$

where $\lambda_i(t) = \lim_{dt \downarrow 0} \mathrm{P}(D_i \in (t, t+dt] | t \leq T_i)/dt$ is the hazard function of the ith individual, $\lambda_0(t)$ is the baseline hazard function, and $\boldsymbol{\beta}$ is the p-dimensional coefficient vector. From model (2), it can be seen that the disease predictors are associated with the hazard function through $r_i(t) = \boldsymbol{\beta}^\mathsf{T} \mathbf{m}_i(t)$. For this reason, $r_i(t)$ is defined to be the *disease risk function* of the ith individual, which is the quantity that we are interested in monitoring later for disease early detection. Regarding $r_i(t)$, it can be checked that its mean among all non-diseased people who are at risk at time t is $\mu_r^{(0)}(t) = \mathrm{E}[r_i(t) | t \leq T_i, \Delta_i = 0] = \boldsymbol{\beta}^\mathsf{T} \boldsymbol{\mu}^{(0)}(t)$, and its mean among all diseased people who are at risk at time t is $\mu_r^{(1)}(t) = \mathrm{E}[r_i(t) | t \leq T_i, \Delta_i = 1] = \boldsymbol{\beta}^\mathsf{T} \boldsymbol{\mu}^{(1)}(t)$, where $\boldsymbol{\mu}^{(1)}(t) = \boldsymbol{\mu}^{(0)}(t) + \boldsymbol{\delta}(t)$.

Model estimation. Next, we discuss estimation of the joint models (1) and (2). To this end, we develop a multivariate local polynomial mixed-effects model estimation procedure, extended from the univariate version discussed in Wu and Zhang (2002). At a given time point $t \in [0, \mathcal{T}]$, consider a small neighborhood $[t-h, t+h]$, where h is a bandwidth parameter. In that neighborhood, $\boldsymbol{\mu}^{(0)}(s)$, $\boldsymbol{\delta}(s)$ and $\mathbf{v}_i(s)$, for $s \in [t-h, t+h]$, can be approximated by their rth-order Taylor's

expansions

$$\boldsymbol{\mu}^{(0)}(s) \approx \sum_{k=0}^{r}(s-t)^k \frac{1}{k!}\frac{\mathrm{d}^k\boldsymbol{\mu}^{(0)}(t)}{\mathrm{d}t^k} = \mathbf{X}(s-t)\begin{bmatrix} \boldsymbol{\mu}^{(0)}(t) \\ \vdots \\ \frac{1}{r!}\frac{\mathrm{d}^r\boldsymbol{\mu}^{(0)}(t)}{\mathrm{d}t^r} \end{bmatrix},$$

$$\boldsymbol{\delta}(s) \approx \sum_{k=0}^{r}(s-t)^k \frac{1}{k!}\frac{\mathrm{d}^k\boldsymbol{\delta}(t)}{\mathrm{d}t^k} = \mathbf{X}(s-t)\begin{bmatrix} \boldsymbol{\delta}(t) \\ \vdots \\ \frac{1}{r!}\frac{\mathrm{d}^r\boldsymbol{\delta}(t)}{\mathrm{d}t^r} \end{bmatrix},$$

$$\mathbf{v}_i(s) \approx \sum_{k=0}^{r}(s-t)^k \frac{1}{k!}\frac{\mathrm{d}^k\mathbf{v}_i(t)}{\mathrm{d}t^k} = \mathbf{X}(s-t)\begin{bmatrix} \mathbf{v}_i(t) \\ \vdots \\ \frac{1}{r!}\frac{\mathrm{d}^r\mathbf{v}_i(t)}{\mathrm{d}t^r} \end{bmatrix},$$

where $\mathbf{X}(t) = \big[\mathbf{I}_{p\times p}, t\mathbf{I}_{p\times p}, \ldots, t^r\mathbf{I}_{p\times p}\big]$, and $\mathbf{I}_{p\times p}$ is the $p\times p$ identity matrix. By these local function approximations, the latent trajectories $\mathbf{m}_i(s)$ in model (2) can be approximated by $\mathbf{X}(s-t)\big[\mathbf{c}(t)+\mathbf{d}(t)+\mathbf{a}_i(t)\big]$, for $s\in[t-h,t+h]$, where $\mathbf{c}(t) = \big[\boldsymbol{\mu}^{(0)}(t)^\mathsf{T}, \ldots, \frac{1}{r!}\big(\frac{\mathrm{d}^r\boldsymbol{\mu}^{(0)}(t)}{\mathrm{d}t^r}\big)^\mathsf{T}\big]^\mathsf{T}$, $\mathbf{d}(t) = \big[\boldsymbol{\delta}(t)^\mathsf{T}, \ldots, \frac{1}{r!}\big(\frac{\mathrm{d}^r\boldsymbol{\delta}(t)}{\mathrm{d}t^r}\big)^\mathsf{T}\big]^\mathsf{T}$, and $\mathbf{a}_i(t) = \big[\mathbf{v}_i(t)^\mathsf{T}, \ldots, \frac{1}{r!}\big(\frac{\mathrm{d}^r\mathbf{v}_i(t)}{\mathrm{d}t^r}\big)^\mathsf{T}\big]^\mathsf{T}$. Consequently, model (1) can be approximated by the following linear mixed-effects model:

$$y_{ik}(t_{ikj}) = \mathbf{e}_k^\mathsf{T}\mathbf{X}(t_{ikj}-t)\big[\mathbf{c}(t)+\Delta_i\mathbf{d}(t)+\mathbf{a}_i(t)\big] + \epsilon_{ik}(t_{ikj}), \qquad \text{for } t_{ikj}\in[t-h,t+h], \tag{3}$$

where $\mathbf{e}_k$ is a p-dimensional vector with its kth element being 1 and all other elements being 0, $\mathbf{c}(t)+\Delta_i\mathbf{d}(t)$ is the fixed-effects term, and $\mathbf{a}_i(t)$ is the random-effects term with mean $\mathbf{0}$ and covariance $\boldsymbol{\Sigma}_b(t)$. To simplify the notation, we will denote the quantities in the brackets by $\mathbf{b}_i(t) = \mathbf{c}(t)+\Delta_i\mathbf{d}(t)+\mathbf{a}_i(t)$. Then, $\mathbf{X}(s-t)\mathbf{b}_i(t)$ is an approximation of $\mathbf{m}_i(s)$, for $s\in[t-h,t+h]$, and the mean and covariance of $\mathbf{b}_i(t)$ are $\mathbf{0}$ and $\boldsymbol{\Sigma}_b(t)$, respectively.

To provide local estimates of $\mathbf{m}_i(s)$ and other quantities in the model (3), let $K(s)$ be a symmetric kernel function with the support $[-1,1]$ and $K(0)=1$. In all numerical examples in this paper, $K(s)$ is chosen to be the Epanechnikov kernel function $K(u) = (1-u^2)I(|u|\le 1)$, which is a symmetric, uni-modal and smooth function with the compact support $[-1,1]$. The Epanechnikov kernel function satisfies the conditions listed in You and Qiu (2021) that are needed for justifying the consistency of parameter estimates. Its simple analytical form can also reduce the computational complexity. Next, we describe a two-stage procedure for estimating the time-varying parameters $\mathbf{c}(t)$, $\mathbf{d}(t)$, $\boldsymbol{\Sigma}_b(t)$ and $\sigma_k^2(t)$, and the time-independent parameter vector $\boldsymbol{\beta}$. To this end, we first derive the local likelihood function for estimating these parameters. It should be pointed out that

although the derivation below is based on the assumption that the longitudinal observations and the random effects have normal distributions, this normality assumption is actually unnecessary to have the consistency of the estimates, as shown in You and Qiu (2021). Assume that the ith individual is in $R(t)$ (i.e., it is at risk at the time t). In the neighborhood $[t-h, t+h]$, if it is assumed that $y_{ik}(t_{ikj})$ follows a normal distribution with mean $\mathbf{e}_k^{\mathsf{T}}\mathbf{X}(t_{ikj}-t)\mathbf{b}_i(t)$ and variance $\sigma_k^2(t)$, then conditional on $\mathbf{b}_i(t)$, the log local-weighted probability density of $\mathcal{Y}_i = \{y_{ik}(t_{ikj}) : k = 1, \ldots, p, j = 1, \ldots, n_{ik}\}$ at t is

$$\begin{aligned}\log f_t(\mathcal{Y}_i|\mathbf{b}_i(t)) &= -\frac{1}{2}\sum_{k=1}^{p}\sum_{j=1}^{n_{ik}}\log\{2\pi\sigma_k^2(t)\}K_h(t_{ikj}-t) \\ &\quad -\frac{1}{2}\sum_{k=1}^{p}\sum_{j=1}^{n_{ik}}\frac{\left[y_{ik}(t_{ikj})-\mathbf{e}_k^{\mathsf{T}}\mathbf{X}(t_{ikj}-t)\mathbf{b}_i(t)\right]^2}{\sigma_k^2(t)}K_h(t_{ikj}-t), \qquad (4)\end{aligned}$$

where $K_h(s) = K(s/h)$. The quantity $-2\log f_t(\mathcal{Y}_i|\mathbf{b}_i(t))$ can also be regarded as a penalized local-weighted least square objective function with the penalty term $\sum_{k=1}^{p}\sum_{j=1}^{n_{ik}}\log\{2\pi\sigma_k^2(t)\}K_h(t_{ikj}-t)$. Thus, the normality assumption mentioned above is not essential for the suggested model estimation procedure. It is used mainly for the convenience in deriving the objective function. Similarly, the log probability density function of the random-effects term $\mathbf{b}_i(t)$ is

$$\log f_t(\mathbf{b}_i(t)) = -\frac{1}{2}\log\det(2\pi\mathbf{\Sigma}_b(t)) - \frac{1}{2}[\mathbf{b}_i(t)-\mathbf{c}(t)-\Delta_i\mathbf{d}(t)]^{\mathsf{T}}\mathbf{\Sigma}_b(t)^{-1}[\mathbf{b}_i(t)-\mathbf{c}(t)-\Delta_i\mathbf{d}(t)].$$

Then, for a given $t \in [0, \mathcal{T}]$, the time-varying parameters $\mathbf{c}(t)$, $\mathbf{d}(t)$, $\mathbf{\Sigma}_b(t)$ and $\sigma_k^2(t)$ can be estimated by maximizing the following local likelihood function:

$$L(\mathbf{c}(t), \mathbf{\Sigma}_b(t), \sigma_k^2(t)) = \prod_{i:t\le T_i}\int f_t(\mathcal{Y}_i|\mathbf{b}_i(t))f_t(\mathbf{b}_i(t))\,d\mathbf{b}_i(t).$$

To solve the above local maximum likelihood estimation problem, we suggest using a local version of the EM algorithm. Similar to the conventional EM algorithm, the local version proceeds by iterating between the expectation step and the maximization step. In the expectation step, the expectation of the log-likelihood is evaluated conditional on the observed data. Then, in the maximization step, parameter estimates are updated by maximizing the conditional expectation of the log-likelihood. Different from the conventional EM algorithm, the current local EM algorithm works with the local likelihood and thus the conditional expectation in the EM algorithm should be taken with respect to the local probability density function. To save space, the local EM algorithm is described in detail in Appendix A.1.

After the time-varying parameters are estimated using the local EM algorithm, we can proceed to the second stage to estimate the time-independent coefficient vector $\boldsymbol{\beta}$ in the survival model (2). Let $\widehat{\mathbf{c}}(t)$, $\widehat{\mathbf{d}}(t)$, $\widehat{\boldsymbol{\Sigma}}_b(t)$, and $\widehat{\sigma}_k^2(t)$ be the final local maximum likelihood estimators of $\mathbf{c}(t)$, $\mathbf{d}(t)$, $\boldsymbol{\Sigma}_b(t)$ and $\sigma_k^2(t)$. Then, $\boldsymbol{\mu}^{(0)}(t)$, $\boldsymbol{\delta}(t)$ and $\boldsymbol{\Sigma}_m(t) = \mathrm{Var}(\mathbf{m}_i(t))$ can be estimated by $\mathbf{X}(0)\widehat{\mathbf{c}}(t)$, $\mathbf{X}(0)\widehat{\mathbf{d}}(t)$ and $\mathbf{X}(0)\widehat{\boldsymbol{\Sigma}}_b(t)\mathbf{X}(0)^\mathsf{T}$, respectively. The random effects term $\mathbf{b}_i(t)$ can be estimated by its corresponding best linear unbiased predictor (BLUP), defined to be $\widehat{\mathbf{b}}_i(t) = \widehat{\mathrm{E}}[\mathbf{b}_i(t)|\mathcal{Y}_i] = \widehat{\mathbf{m}}_{b,i}(t)$, where $\widehat{\mathbf{m}}_{b,i}(t)$ is defined to be $\mathbf{m}_{b,i}(t)$ given in Appendix A.1, after certain unknown quantities are replaced by their local maximum likelihood estimators. The quantity $\mathbf{m}_i(t)$ can be estimated similarly by $\mathbf{X}(0)\widehat{\mathbf{b}}_i(t)$. To estimate $\boldsymbol{\beta}$, we can first plug in the estimated values of $\mathbf{m}_i(t)$ into the following Cox partial likelihood function:

$$\mathrm{pl}(\boldsymbol{\beta}) = \sum_{i=1}^{m} \Delta_i \left[\boldsymbol{\beta}^\mathsf{T} \mathbf{m}_i(T_i) - \log \left\{ \sum_{l=1}^{m} \exp\{\boldsymbol{\beta}^\mathsf{T} \mathbf{m}_l(T_i)\} I(T_l \geq T_i) \right\} \right]. \tag{5}$$

Then, $\boldsymbol{\beta}$ can be estimated by the maximizer of (5), denoted as $\widehat{\boldsymbol{\beta}}$, which can be computed by the standard Newton-Raphson algorithm.

In the local EM algorithm described in Appendix A.1, the bandwidth parameter h needs to be selected properly. To this end, we suggest using a multivariate version of the leave-one-point-out cross-validation (PTCV) criterion (see Altman 1990; Wu and Zhang 2002), described below. Let $\tau_1 < \ldots < \tau_G$ be all distinct time points in the set $\{t_{ikj} : i = 1, \ldots, m, k = 1, \ldots, p, j = 1, \ldots, n_{ik}\}$. Define $\widehat{m}_{ik}^{(-g)}(t)$ to be the estimate of $m_{ik}(t)$ when all observations at τ_g are excluded. Then, the PTCV score is defined as

$$\mathrm{PTCV}(h) = \sum_{g=1}^{G} \sum_{i,k,j:t_{ikj}=\tau_g} \left[y_{ik}(t_{ikj}) - \widehat{m}_{ik}^{(-g)}(t_{ikj}) \right]^2,$$

which measures the difference between the predicted values of $m_{ik}(t_{ikj})$ and the actual observations at t_{ikj}. The bandwidth h is then selected by minimizing the PTCV score $\mathrm{PTCV}(h)$. In this bandwidth selection procedure, in cases when the scales of y_{ik}, for different k, are very different, we recommend re-scaling them in advance so that different disease predictors have similar sample variances.

3 Dynamic Disease Screening

To detect a disease in concern early for a new individual not contained in the training data, we need to sequentially monitor his/her longitudinal pattern of the observed disease predictors. Assume that the kth disease predictor $y_k^*(t)$ is observed at times $t_{k1}^*, t_{k2}^*, \ldots$, for $k = 1, 2, \ldots, p$, where the superscript "*" is used in the notations here to distinguish them from those in the training data described in model (1). For these longitudinal data, it is assumed that they follow the model

$$y_k^*(t_{kj}^*) = m_k^*(t_{kj}^*) + \epsilon_k^*(t_{kj}^*), \qquad \text{for } k = 1, 2, \ldots, p,$$

where $m_k^*(t)$ is the latent longitudinal trajectory of the kth disease predictor, and $\epsilon_k^*(t)$ is the random error term. If the given individual is a non-diseased person, then his/her disease risk at time t can still be quantified by $\widehat{r}^*(t) = \widehat{\boldsymbol{\beta}}^{\mathsf{T}}\mathbf{m}^*(t)$, where $\widehat{\boldsymbol{\beta}}$ is obtained by the local EM algorithm from the training dataset, as discussed in Section 2, and $\mathbf{m}^*(t) = (m_1^*(t), \ldots, m_p^*(t))^{\mathsf{T}}$.

Assume that the current time point is t during the sequential monitoring for the new individual. Then, given all the history data $\{(t_{kj}^*, y_k^*(t_{kj}^*)) : t_{kj}^* \le t\}$ of his/her observed disease predictors, we would like to make a judgment whether the new individual has a significantly higher disease risk or not at time t, compared to the non-diseased people in the training data. The related hypotheses are $H_0 : \mathrm{E}[\widehat{r}^*(t)] = \mu_r^{(0)}(t)$ versus $H_1 : \mathrm{E}[\widehat{r}^*(t)] > \mu_r^{(0)}(t)$. By the SPC terminology, the new individual is declared to be out-of-control (OC) at time t if the null hypothesis is rejected. To assess the hypotheses, we consider using the following test statistic:

$$U(t) = \frac{\widehat{r}^*(t) - \mu_r^{(0)}(t)}{\sqrt{\widehat{\boldsymbol{\beta}}^{\mathsf{T}}\widehat{\boldsymbol{\Sigma}}_m(t)\widehat{\boldsymbol{\beta}}}} = \frac{\widehat{\boldsymbol{\beta}}^{\mathsf{T}}\mathbf{m}^*(t) - \mu_r^{(0)}(t)}{\sqrt{\widehat{\boldsymbol{\beta}}^{\mathsf{T}}\widehat{\boldsymbol{\Sigma}}_m(t)\widehat{\boldsymbol{\beta}}}},$$

where $\widehat{\boldsymbol{\Sigma}}_m(t)$ is obtained by the joint modeling approach discussed in Section 2. To estimate the latent trajectory $\mathbf{m}^*(t)$ of the new individual, let us consider the following objective function:

$$\begin{aligned} L(\mathbf{m}^*(t)) = &-\frac{1}{2}\sum_{k=1}^{p}\sum_{j:t_{kj}^*\le t}\frac{\left[y_k^*(t_{kj}^*) - \mathbf{e}_k^{\mathsf{T}}\mathbf{m}^*(t)\right]^2}{\sigma_k^2(t)}(1-\lambda)^{(t-t_{kj}^*)/\bar{d}} \\ &- \frac{1}{2}[\mathbf{m}^*(t) - \boldsymbol{\mu}^{(0)}(t)]^{\mathsf{T}}\boldsymbol{\Sigma}_m(t)^{-1}[\mathbf{m}^*(t) - \boldsymbol{\mu}^{(0)}(t)], \end{aligned}$$

where $\bar{d}$ is the average distance between two consecutive observation times that can be estimated from the training dataset, and λ is a weighting parameter. It can be seen that the expression of $L(\mathbf{m}^*(t))$ is similar to that of (4), with $\mathbf{X}(t_{ikj} - t)$ being replaced by $\mathbf{X}(0)$ and $K_h(t_{ikj} -$

t) by $(1-\lambda)^{(t-t^*_{kj})/\bar{d}}$. Thus, $L(\mathbf{m}^*(t))$ can be regarded as a local weighted likelihood with an exponential weighting function. In $L(\mathbf{m}^*(t))$, past observations have been used and receive the weight $(1-\lambda)^{(t-t^*_{kj})/\bar{d}}$, which decreases exponentially fast as $t - t^*_{kj}$ increases. Namely, the older the observations, the exponentially smaller the weights they receive. The estimate of $\mathbf{m}^*(t)$ is then defined to be the maximizer of $L(\mathbf{m}^*(t))$. Since $L(\mathbf{m}^*(t))$ has a quadratic form of $\mathbf{m}^*(t)$, the estimator for $\mathbf{m}^*(t)$ can be derived to be

$$\widehat{\mathbf{m}}^*(t) = \Bigg[\sum_{k=1}^{p}\sum_{j:t^*_{kj}\le t}\frac{\mathbf{e}_k\mathbf{e}_k^{\mathsf{T}}}{\widehat{\sigma}_k^2(t)}(1-\lambda)^{(t-t^*_{kj})/\bar{d}} + \widehat{\mathbf{\Sigma}}_m(t)^{-1}\Bigg]^{-1} \times \Bigg[\sum_{k=1}^{p}\sum_{j:t^*_{kj}\le t}\frac{y_k^*(t^*_{kj})\mathbf{e}_k}{\widehat{\sigma}_k^2(t)}(1-\lambda)^{(t-t^*_{kj})/\bar{d}} + \widehat{\mathbf{\Sigma}}_m(t)^{-1}\widehat{\boldsymbol{\mu}}^{(0)}(t)\Bigg],$$

where $\mathbf{\Sigma}_m(t)$, $\sigma_k^2(t)$ and $\boldsymbol{\mu}^{(0)}(t)$ have been replaced by $\widehat{\mathbf{\Sigma}}_m(t)$, $\widehat{\sigma}_k^2(t)$ and $\widehat{\boldsymbol{\mu}}^{(0)}(t)$, respectively, which are obtained from the training dataset as discussed in Section 2. Then, a natural charting statistic is

$$\widehat{U}(t) = \frac{\widehat{\boldsymbol{\beta}}^{\mathsf{T}}\widehat{\mathbf{m}}^*(t) - \widehat{\mu}_r^{(0)}(t)}{\sqrt{\widehat{\boldsymbol{\beta}}^{\mathsf{T}}\widehat{\mathbf{\Sigma}}_m(t)\widehat{\boldsymbol{\beta}}}}, \tag{6}$$

and the chart gives a signal at time t if $\widehat{U}(t) > \rho$, where $\rho > 0$ is a control limit.

In the control chart (6), the weighting parameter λ controls the amount of history information used in process monitoring at the current time point. If λ is chosen larger, then more weight is assigned to the current observation and less weight is assigned to the previous observations. Next, we describe a method to determine λ based on the training data. For a given value of λ, consider the following loss function

$$\widetilde{L}(\mathbf{m}_i(t)) = -\frac{1}{2}\sum_{k=1}^{p}\sum_{j:t_{ikj}<t}^{n_{ik}}[y_{ik}(t_{ikj}) - \mathbf{e}_k^{\mathsf{T}}\mathbf{m}_i(t)]^2(1-\lambda)^{(t-t_{ikj})/\bar{d}} - \frac{1}{2}[\mathbf{m}_i(t) - \boldsymbol{\mu}^{(0)}(t)]^{\mathsf{T}}\mathbf{\Sigma}_m(t)^{-1}[\mathbf{m}_i(t) - \boldsymbol{\mu}^{(0)}(t)].$$

It can be seen that $\widetilde{L}(\mathbf{m}_i(t))$ is the same as $L(\mathbf{m}_i(t))$, except that observations at the current time point t are excluded when computing the former quantity. Let $\widetilde{\mathbf{m}}_{i,\lambda}(t)$ be the minimizer of $\widetilde{L}(\mathbf{m}_i(t))$. Then, $\widetilde{\mathbf{m}}_{i,\lambda}(t)$ can be regarded as a prediction of $\mathbf{m}_i(t)$ based on the history data, and the prediction error can be defined by

$$\mathrm{PE}(\lambda) = \sum_{i=1}^{m}\sum_{k=1}^{p}\sum_{j=1}^{n_{ik}}[y_{ik}(t_{ikj}) - \mathbf{e}_k^{\mathsf{T}}\widetilde{\mathbf{m}}_{i,\lambda}(t_{ikj})]^2.$$

Then, the value of λ can be chosen by minimizing $\text{PE}(\lambda)$.

In the dynamic disease screening literature (e.g., Qiu and Xiang, 2014), the performance of a control chart like (6) is usually evaluated by the average time to signal (ATS). Let $\Delta^*_{\text{signal}} = 1$ denote the individual under monitoring receives a signal and 0 otherwise, and T^*_{signal} denote the time when the individual receives the signal. Then, ATS is defined to be $\text{ATS} = \text{E}[T^*_{\text{signal}}|\Delta^*_{\text{signal}} = 1]$. The IC value of ATS, denoted as ATS_0, is the average time to signals when the processes under monitoring are IC, and the OC value of ATS, denoted as ATS_1, is the average time to signals when the processes under monitoring are OC. When comparing different control charts, we usually fix ATS_0 at a given level, and then compare their ATS_1 values. Control charts with smaller ATS_1 values are considered better because they can provide signals to OC individuals sooner.

The control limit ρ of the chart (6) is usually selected such that a pre-specified value of ATS_0 is attained. To determine the value of ρ, the block bootstrap method discussed in Qiu and Xiang (2014) can be used. By this method, the original training dataset is partitioned into two parts, one of which is used for model estimation and the other one is used for determining the control limit. Then, individuals in the second part are re-sampled with replacement for online monitoring, and a numerical searching algorithm is used for determining the value of ρ.

4 Some Extensions

The proposed method discussed in the previous sections can be extended in several different directions, which are briefly discussed in this section.

4.1 Cases with longitudinal categorical or count data

In Section 2, we focus on cases when the disease predictors are continuous variables. Here, we comment that the proposed method can be extended to cases when the disease predictors take categorical or count values, by using the generalized linear modeling framework. To this end, consider the following generalized linear model that is modifed from Model (1):

$$\begin{aligned}
\text{E}[y_{ik}(t_{ikj})|\Delta_i, v_{ik}(t_{ikj}), t_{ikj} \le T_i] &= g_k^{-1}\big(\mu_k^{(0)}(t_{ikj}) + \Delta_i\delta_k(t_{ikj}) + v_{ik}(t_{ikj})\big), \\
\text{Var}(y_{ik}(t_{ikj})|\Delta_i, v_{ik}(t_{ikj}), t_{ikj} \le T_i) &= \phi_k V_k\big(\text{E}[y_{ik}(t_{ikj})|\Delta_i, v_{ik}(t_{ikj}), t_{ikj} \le T_i]\big),
\end{aligned}$$

where $g_k(\cdot)$, $V_k(\cdot)$, and ϕ_k are the canonical link function, variance function, and dispersion parameter for the kth disease predictor, respectively, for $k = 1, 2, \ldots, p$. Depending on the type of values (e.g., continuous, binary, or count) of the kth disease predictor, one can choose $g_k(\cdot)$, $V_k(\cdot)$, and ϕ_k accordingly (cf., McCullagh and Nelder, 2019). Model (1) can be regarded as a special case of the above model when $g_k(x) = x$, $V_k(x) = 1$, and $\phi_k = \sigma_k^2$. When the kth disease predictor is categorical with N_k categories $a_1, \ldots, a_{N_k}$, the following generalized linear model can be considered: for $l = 2, \ldots, N_k$,

$$\log\left[\frac{\mathrm{P}(y_{ik}(t_{ikj}) = a_l|\Delta_i, v_{ik}(t_{ikj}), t_{ikj} \le T_i)}{\mathrm{P}(y_{ik}(t_{ikj}) = a_1|\Delta_i, v_{ik}(t_{ikj}), t_{ikj} \le T_i)}\right] = \mu_{kl}^{(0)}(t_{ikj}) + \Delta_i\delta_{kl}(t_{ikj}) + v_{ikl}(t_{ikj}),$$

where a_1 is the reference category. When the kth disease predictor is an ordinal categorical variable that takes the integer values $1, \ldots, N_k$, the following proportional odds model can be considered:

$$\log\left[\frac{P(y_{ik}(t_{ikj}) \le l|\Delta_i, v_{ik}(t_{ikj}), t_{ikj} \le T_i)}{P(y_{ik}(t_{ikj}) > l|\Delta_i, v_{ik}(t_{ikj}), t_{ikj} \le T_i)}\right] = \mu_{kl}^{(0)}(t_{ikj}) + \Delta_i\delta_k(t_{ikj}) + v_{ik}(t_{ikj}),$$

for $l = 1, \ldots, N_k - 1$. The parameters in the above models can be estimated by the local EM algorithm similarly to the one discussed in Section 2. However, there are generally no closed-form expressions for the integrals involved in the related expectations. So, more sophisticated techniques might be required for parameter estimation. A systematic research on this method extension is deferred to our future research.

4.2 Accommodation of covariates

Our proposed method discussed in the previous sections can be extended to include some longitudinal covariates for describing their impact on the latent longitudinal trajectories of the disease predictors. Let $\mathbf{x}_{ik}(t)$ be the longitudinal covariates for the kth disease predictor. Then, Model (1) can be generalized to

$$y_{ik}(t_{ikj}) = \mu_k^{(0)}(t_{ikj}) + \Delta_i\delta_k(t_{ikj}) + v_{ik}(t_{ikj}) + \boldsymbol{\gamma}_k^{\mathsf{T}}\mathbf{x}_{ik}(t_{ikj}) + \epsilon_{ik}(t_{ikj}),$$

where $\{\boldsymbol{\gamma}_k\}$ are coefficient vectors. This model can still be estimated by the local EM algorithm discussed in Section 2. Then, the part $\mu_k^{(0)}(t_{ikj}) + \Delta_i\delta_k(t_{ikj}) + v_{ik}(t_{ikj}) + \boldsymbol{\gamma}_k^{\mathsf{T}}\mathbf{x}_{ik}(t_{ikj})$ in the above model can replace $m_{ik}(t_{ikj})$ in model (2), and the dynamic screening method can be constructed in the same way as that discussed in Section 3 afterwards. By using the above longitudinal model, the covariate effect would be assumed to be multiplicative with respect to the hazard rate in model

(2). This assumption could be violated in certain applications. In such cases, some more flexible alternatives can be considered, which include the time-varying coefficient model discussed in Section 4.3 below and the nonparametric regression model discussed in Section 7.

4.3 Survival model with time-varying coefficients

The proportional hazards model (2) can also be generalized to the following time-varying coefficient model to allow the impact of the longitudinal predictors on the survival outcomes to change over time:

$$\lambda_i(t) = \lambda_0(t) \exp\{\boldsymbol{\beta}(t)^{\mathsf{T}} \mathbf{m}_i(t)\}, \tag{7}$$

where $\boldsymbol{\beta}(t)$ is the time-varying coefficient vector. Many different approaches (e.g., regression splines) for estimation of Model (7) have been proposed in the literature. To be consistent with the model estimation method discussed in the previous sections, a kernel smoothing method that is similar to the ones discussed in Tian et al. (2005) and You and Qiu (2021) can be considered here. More specifically, $\boldsymbol{\beta}(t)$ can be estimated by

$$\widehat{\boldsymbol{\beta}}(t) = \text{argmax}_{\boldsymbol{\beta}} \, \text{pl}(\boldsymbol{\beta}, t),$$

where

$$\text{pl}(\boldsymbol{\beta}, t) = \sum_{i=1}^{m} K_h(T_i - t) \Delta_i \left[\boldsymbol{\beta}^{\mathsf{T}} \mathbf{m}_i(T_i) - \log \left\{ \sum_{l=1}^{m} \exp\{\boldsymbol{\beta}^{\mathsf{T}} \mathbf{m}_l(T_i)\} I(T_l \geq T_i) \right\} \right].$$

In the above maximization procedure, $\mathbf{m}_l(t)$ needs to be estimated, which can be accomplished using the local EM algorithm discussed in Section 2 and Appendix A.

After $\widehat{\boldsymbol{\beta}}(t)$ is obtained, Expression (6) for defining the charting statistic for process monitoring can be changed to

$$\widehat{U}(t) = \frac{\widehat{\boldsymbol{\beta}}(t)^{\mathsf{T}} \widehat{\mathbf{m}}^*(t) - \widehat{\mu}_r^{(0)}(t)}{\sqrt{\widehat{\boldsymbol{\beta}}(t)^{\mathsf{T}} \widehat{\boldsymbol{\Sigma}}_m(t) \widehat{\boldsymbol{\beta}}(t)}}. \tag{8}$$

5 Simulation Study

In this section, a simulation study is conducted to assess the numerical performance of the proposed method. In each simulation example, the training dataset contains observations of $m = 500$ individuals with $p = 4$ disease predictors. The basic time unit is assumed to be $\omega = 0.001$, and the

study time period is $[0, 2]$. Then, all observation times would be integer multiples of ω. For instance, the observation times of an individual could be $\{0.1, 0.15, 0.531, 0.734, 0.85\}$, all of which are integer multiples of 0.001. But, 0.1542 would be rounded to 0.154 in the simulation study. The following five different cases are considered. In cases (I) and (II), longitudinal observations of the disease predictors are generated from the following mixed-effects model: for $i = 1, \ldots, m$ and $k = 1, \ldots, p$,

$$\begin{aligned} y_{ik}(t) &= m_{ik}(t) + \epsilon_{ik}(t) && (9) \\ &= 2 + \sin(\pi t) + \xi_{ik1}\phi_1(t) + \xi_{ik2}\phi_2(t) + \xi_{ik3}\phi_3(t) + \epsilon_{ik}(t), && (10) \end{aligned}$$

where $\phi_1(t) = (t-1)^2$, $\phi_2(t) = \sin(2\pi t)$, $\phi_3(t) = \cos(2\pi t)$, $\{\xi_{ikl}, l = 1, 2, 3\}$ are iid random numbers generated from the $N(0, 1)$ distribution. The random errors $\{\epsilon_{ik}(t)\}$ are generated i.i.d. from the $N(0, 1)$ distribution and the t-distribution with 3 degrees of freedom, respectively, in cases (I) and (II). In the above model, $\xi_{ik1}\phi_1(t) + \xi_{ik2}\phi_2(t) + \xi_{ik3}\phi_3(t)$ is the time-dependent random-effects term for describing within-subject and within-disease-predictor data correlation. For simplicity, it is assumed that different disease predictors are observed at the same observation times for a given individual. More precisely, we let $t_{i1j} = \cdots = t_{ipj}$, for all i and j, and t_{i1j} is generated from the uniform distribution in the interval $[(j-1)/200, j/200]$, for each i and j. This restriction is due to the fact that the existing DySS methods that will be compared with the proposed method cannot handle cases when different disease predictors are observed at different time points. Note that all observation times t_{ikj} will end at the event time T_i determined by the survival model specified below, for all i, j and k. To study the impact of the percentage of diseased people in the training dataset on the performance of the proposed method, we consider three scenarios when the baseline hazard function is $\lambda_0(t) = 0.03$, $\lambda_0(t) = 0.02$ and $\lambda_0(t) = 0.01$, respectively, so that the percentages of diseased people in the second and third scenarios are approximately 2/3 and 1/3 of the one in the first scenario. The survival outcomes are generated from the model (2). In cases (III)–(V), longitudinal observations of the disease predictors are generated from the following model:

$$\begin{aligned} y_{ik}(t) &= m_{ik}(t) + \epsilon_{ik}(t) && (11) \\ &= 2 + \sin(\pi t/2) + g_{ik}(t) + \epsilon_{ik}(t), && (12) \end{aligned}$$

where $\{g_{ik}(t)\}$ are i.i.d. Gaussian processes with the covariance function $\sigma(s, t) = 2\exp\{-10(s-t)^2\}$, and $\{\epsilon_{ik}(t)\}$ follow the zero mean normal distribution with the standard deviation being 1, 0.5, and 0, respectively. In cases (III)–(V), the proportional hazards model (2) is assumed to be valid, and the baseline hazard function is assumed to be $\lambda_0(t) = 0.03$, $\lambda_0(t) = 0.02$ or $\lambda_0(t) = 0.01$. In cases

(I)–(V), the proportional hazards assumption is assumed valid. Next, we consider several cases when this assumption is violated or the covariance function $\sigma(s,t)$ has a more complex form. More specifically, in case (VI), we assume that the longitudinal observations are generated in the same way as that in case (I) (i.e., from Models (9–10)), but the survival outcomes are generated from the following non-proportional hazards model:

$$\lambda_i(t) = \lambda_0(t)\{1 + \boldsymbol{\beta}^{\mathsf{T}}\mathbf{m}_i(t)\}^2. \tag{13}$$

In cases (VII) and (VIII), it is assumed that the longitudinal observations are generated from Models (11–12) with the following rational-quadratic variance-covariance function:

$$\sigma(s,t) = \{1 + 2(s-t)^2\}^{-0.5}, \tag{14}$$

and the random errors $\{\epsilon_{ik}(t)\}$ follow the $N(0, 2^2)$ distribution. The survival outcomes are generated from the proporional hazards model (2) in case (VII), and the non-proportional hazards model (13) in case (VIII). In cases (VI–VIII), the baseline hazard function is $\lambda_0(t) = 0.03$, $\lambda_0(t) = 0.02$, and $\lambda_0(t) = 0.01$ respectively. In all eight cases described above, simulations are repeated for 500 times, the true regression coefficients are $\boldsymbol{\beta} = (0.5, 0.4, -0.3, -0.2)^{\mathsf{T}}$, and the true censoring times are $C_i = \mathcal{T}$, for each i. The control limit of the proposed chart is selected using the block bootstrap method discussed at the end of Section 3 such that $ATS_0 = 370 \times 0.001$.

The proposed method is then compared with two existing methods on multivariate dynamic disease screening suggested by Qiu and Xiang (2015) and You and Qiu (2020). The method by Qiu and Xiang (2015) has two versions: the one using a multivariate EWMA chart and the one combining multiple univariate control charts. These two versions are denoted as "DySS-M" and "DySS-C", respectively, where "M" represents "multivariate" and "C" represents "combination of multiple univariate charts." The risk monitoring method by You and Qiu (2020) is denoted as "Risk-Mnt", and the method proposed in this paper is denoted as "New". Among the four methods, "DySS-M" and "DySS-C" use the longitudinal observations of the non-diseased people in the training dataset only for estimating the regular longitudinal patterns of the disease predictors. "Risk-Mnt" first uses a proportional hazards model to quantify the relationship between the disease predictors and the survival outcomes, and then monitors a linear combination of the disease predictors for disease screening of individual people. Its major difference from the proposed method "New" is that it did not use a joint modeling framework when building the functional relationship between the disease predictors and the survival outcomes. In Table 1, we first compare the

estimated regression coefficients in the survival model (2) by "Risk-Mnt" and "New" in terms of the bias and the overall mean squared error (MSE). Because the proportional hazards assumption is violated in Cases (VI) and (VIII) and it is meaningless to compare the estimated regression coefficients with the true regression coefficients of the survival model in such cases, only Cases (I–V) and (VII) are considered here. From the table, it can be seen that in Cases (I–IV) and (VII) when the longitudinal measurements are observed with random errors, the regression coefficients estimated by "Risk-Mnt" have much larger bias and MSE, compared to those by "New". In case (V) when the longitudinal measurements are observed without random errors, the two methods perform similarly.

To evaluate the OC performance of the four methods, we compare their ATS_1 values. To this end, we first simulate 1,000 longitudinal processes from model (1) with $\Delta_i = 0$, and denote the simulated longitudinal processes by $\mathbf{m}_{i,\mathrm{IC}}(t)$. Then, we add shifts to these processes in the direction of $\boldsymbol{\beta}$:

$$\mathbf{m}^{\delta}_{i,\mathrm{OC}}(t) = \mathbf{m}_{i,\mathrm{IC}}(t) + \delta\boldsymbol{\beta},$$

where δ is actually the shift size in disease risks. Then, we evaluate different methods by comparing their ATS_1 values, and the one with a smaller ATS_1 value is considered better. The computed ATS_1 values of the four methods when their ATS_0 values are all set to be 370×0.001 are presented in Figures 1 and 2. From the figures, we can have the following conclusions. First, all methods can reach the nominal ATS_0 value when the shift is 0. Second, "New" has the smallest ATS_1 values in all eight cases, compared to the other three methods for different shift sizes. Because the longitudinal measurements are observed with random errors in these cases, "Risk-Mnt" cannot provide a good estimate of $\boldsymbol{\beta}$ and this explains why it performs worse than "New". In Case (V) when the longitudinal measurements are observed without random errors, "Risk-Mnt" and "New" perform similarly. Third, when the baseline hazard decreases, the number of diseased people in the training dataset would decrease as well, and thus both "Risk-Mnt" and "New" become less efficient. As a comparison, the perfromance of "DySS-M" and "DySS-C" does not change much because they do not rely on the survival information in the training dataset. Fourth, in cases (VI) and (VIII) when the proportional hazards assumption is violated, the proposed method "New" still demonstrates a reasonable performance. Fifth, by comparing cases (III)–(V) when the noise level gets smaller and smaller, it can be seen that all four methods perform better and better, which is intuitively reasonable.

The true positive rates (TPR) and the false positive rates (FPR) of the four methods are presented in Table 2. From the table, it can be seen that in all cases considered here, the two methods that use the survival information in the training dataset (i.e., "Risk-Mnt" and "New") perform better than the two methods that do not use the survival information (i.e., "DySS-M" and "DySS-C"). By comparing "Risk-Mnt" and "New", we can see that they have similar FPR values, but "New" has much larger TPR values in all cases except Case (V) where the TPR values of "New" are slightly larger than those of "Risk-Mnt" since the longitudinal observations are free of noise in such a case. Therefore, the proposed method "New" also demonstrates good performance in terms of TPR and FPR.

We would like to mention that, in the current setup, it needs an average of $n^* = 370 \times 0.001 \times 200 = 74$ longitudinal observations for each non-diseased individual to receive a false signal from the control chart (6). In the SPC literature, a relatively large n^* value is often used to facilitate the evaluation of different methods in terms of ATS, since the signal times for diseased people are often much sooner than those for non-diseased people. In some applications, however, the number of longitudinal observations for each individual could be relatively small. To evaluate the numerical performance of the proposed dynamic disease screening method and the related competing methods in such cases, we also consider the study design in which the observation time t_{i1j} is generated from the uniform distribution in the interval $[(j-1)/100, j/100]$, for each i and j, and all other setups are kept the same as before. In such cases, an average of $n^* = 370 \times 0.001 \times 100 = 37$ longitudinal observations are needed for each non-diseased individual to receive a false signal from the control chart (6). The related results are included in the supplementary material, from which it can be seen that similar conclusions to those made above can be obtained and the proposed method still has an overall better performance compared to its peers.

6 Application to the Primary Biliary Cirrhosis Data

We use a dataset from the Primary Biliary Cirrhosis (PBC) study conducted at the Mayo Clinic as an example to illustrate the application of the proposed method. PBC is a chronic disease when the small bile ducts of the liver become injured, which eventually would lead to cirrhosis of the liver. The dataset contains observed data of 312 PBC patients, among which 172 patients died from PBC during the study and 140 were censored due to either survival during the study or lost to follow-up.

Table 1: Biases and mean squared errors (MSE) of estimated regression coefficients by "Risk-Mnt" and "New" in cases (I)–(V) and (VII). Numbers in parentheses are the standard errors.

Case	$\lambda_0(t)$	Method	Bias($\widehat{\beta}_1$)	Bias($\widehat{\beta}_2$)	Bias($\widehat{\beta}_3$)	Bias($\widehat{\beta}_4$)	Overall MSE
(I)	0.03	Risk-Mnt	-0.2264 (0.0047)	-0.1791 (0.0049)	0.1298 (0.0047)	0.0843 (0.0046)	0.1522 (0.0035)
		New	-0.0040 (0.0070)	0.0064 (0.0073)	-0.0035 (0.0067)	-0.0087 (0.0063)	0.0937 (0.0033)
	0.02	Risk-Mnt	-0.2321 (0.0060)	-0.1782 (0.0058)	0.1370 (0.0055)	0.0964 (0.0056)	0.1802 (0.0049)
		New	-0.0127 (0.0082)	0.0110 (0.0087)	0.0094 (0.0079)	0.0087 (0.0081)	0.1352 (0.0047)
	0.01	Risk-Mnt	-0.2151 (0.0084)	-0.1832 (0.0084)	0.1287 (0.0082)	0.0932 (0.0078)	0.2398 (0.0073)
		New	0.0219 (0.0124)	-0.0016 (0.0118)	-0.0103 (0.0117)	-0.0039 (0.0113)	0.2788 (0.0105)
(II)	0.03	Risk-Mnt	-0.4035 (0.0028)	-0.3189 (0.0027)	0.2321 (0.0025)	0.1562 (0.0027)	0.3571 (0.0038)
		New	-0.0126 (0.0067)	0.0060 (0.0074)	0.0139 (0.0075)	0.0067 (0.0071)	0.0975 (0.0036)
	0.02	Risk-Mnt	-0.4013 (0.0036)	-0.3155 (0.0033)	0.2311 (0.0030)	0.1564 (0.0034)	0.3601 (0.0048)
		New	-0.0015 (0.0085)	0.0042 (0.0087)	-0.0033 (0.0081)	0.0052 (0.0086)	0.1428 (0.0047)
	0.01	Risk-Mnt	-0.4125 (0.0042)	-0.3211 (0.0041)	0.2423 (0.0043)	0.1562 (0.0043)	0.3911 (0.0060)
		New	-0.0136 (0.0121)	-0.0006 (0.0116)	-0.0061 (0.0126)	0.0005 (0.0123)	0.2948 (0.0116)
(III)	0.03	Risk-Mnt	-0.4452 (0.0014)	-0.3537 (0.0013)	0.2688 (0.0014)	0.1781 (0.0014)	0.4312 (0.0019)
		New	0.0328 (0.0058)	0.0352 (0.0062)	-0.0141 (0.0057)	-0.0194 (0.0059)	0.0720 (0.0023)
	0.02	Risk-Mnt	-0.4452 (0.0016)	-0.3582 (0.0015)	0.2673 (0.0016)	0.1796 (0.0016)	0.4351 (0.0020)
		New	0.0321 (0.0071)	0.0166 (0.0071)	-0.0208 (0.0069)	-0.0118 (0.0068)	0.0994 (0.0032)
	0.01	Risk-Mnt	-0.4468 (0.0022)	-0.3567 (0.0023)	0.2676 (0.0023)	0.1767 (0.0023)	0.4399 (0.0031)
		New	0.0232 (0.0101)	0.0411 (0.0103)	-0.0299 (0.0100)	-0.0216 (0.0103)	0.2095 (0.0072)
(IV)	0.03	Risk-Mnt	-0.3363 (0.0025)	-0.2640 (0.0023)	0.2043 (0.0025)	0.1336 (0.0024)	0.2542 (0.0025)
		New	0.0122 (0.0047)	0.0190 (0.0050)	-0.0045 (0.0046)	-0.0108 (0.0048)	0.0459 (0.0015)
	0.02	Risk-Mnt	-0.3372 (0.0027)	-0.2699 (0.0028)	0.2079 (0.0027)	0.1324 (0.0029)	0.2629 (0.0028)
		New	0.0054 (0.0055)	0.0092 (0.0056)	-0.0071 (0.0054)	-0.0114 (0.0057)	0.0614 (0.0020)
	0.01	Risk-Mnt	-0.3385 (0.0038)	-0.2685 (0.0040)	0.2002 (0.0041)	0.1326 (0.0040)	0.2755 (0.0042)
		New	0.0059 (0.0077)	0.0185 (0.0079)	-0.0137 (0.0080)	-0.0066 (0.0079)	0.1250 (0.0043)
(V)	0.03	Risk-Mnt	0.0105 (0.0044)	0.0126 (0.0043)	-0.0091 (0.0040)	0.0010 (0.0044)	0.0370 (0.0012)
		New	0.0022 (0.0043)	0.0088 (0.0043)	-0.0058 (0.0039)	0.0033 (0.0043)	0.0355 (0.0011)
	0.02	Risk-Mnt	0.0122 (0.0055)	0.0127 (0.0052)	-0.0101 (0.0048)	0.0052 (0.0055)	0.0549 (0.0018)
		New	0.0064 (0.0054)	0.0081 (0.0051)	-0.0065 (0.0047)	0.0077 (0.0054)	0.0528 (0.0017)
	0.01	Risk-Mnt	0.0251 (0.0075)	0.0084 (0.0074)	-0.0274 (0.0077)	0.0049 (0.0070)	0.1119 (0.0039)
		New	0.0180 (0.0073)	0.0031 (0.0073)	-0.0228 (0.0076)	0.0075 (0.0069)	0.1070 (0.0037)
(VII)	0.03	Risk-Mnt	-0.3364 (0.0024)	-0.2686 (0.0024)	0.2006 (0.0024)	0.1351 (0.0023)	0.2549 (0.0024)
		New	0.0066 (0.0050)	0.0125 (0.0048)	-0.0078 (0.0046)	0.0017 (0.0044)	0.0446 (0.0015)
	0.02	Risk-Mnt	-0.3375 (0.0028)	-0.2672 (0.0028)	0.1998 (0.0027)	0.1356 (0.0029)	0.2593 (0.0029)
		New	0.0089 (0.0057)	0.0211 (0.0056)	-0.0177 (0.0056)	0.0002 (0.0053)	0.0625 (0.0021)
	0.01	Risk-Mnt	-0.3393 (0.0037)	-0.2648 (0.0043)	0.2009 (0.0040)	0.1327 (0.0039)	0.2739 (0.0039)
		New	0.0005 (0.0077)	0.0237 (0.0081)	-0.0076 (0.0083)	-0.0126 (0.0079)	0.1281 (0.0046)

Table 2: TPR and FPR of the four methods "DySS-M", "DySS-C", "Risk-Mnt" and "New" in Cases (I)–(VIII).

		DySS-M		DySS-C		Risk-Mnt		New	
Case	$\lambda_0(t)$	TPR	FPR	TPR	FPR	TPR	FPR	TPR	FPR
(I)	0.03	0.3139 (0.0034)	0.4531 (0.0019)	0.4212 (0.0036)	0.4232 (0.0016)	0.5609 (0.0039)	0.4026 (0.0015)	0.6394 (0.0035)	0.4021 (0.0016)
	0.02	0.3136 (0.0042)	0.4513 (0.0018)	0.4255 (0.0042)	0.4225 (0.0015)	0.5431 (0.0049)	0.3986 (0.0015)	0.6247 (0.0042)	0.3984 (0.0016)
	0.01	0.3211 (0.0057)	0.4507 (0.0017)	0.4250 (0.0061)	0.4207 (0.0015)	0.5070 (0.0068)	0.3983 (0.0014)	0.5970 (0.0061)	0.3974 (0.0015)
(II)	0.03	0.3132 (0.0035)	0.4790 (0.0017)	0.4067 (0.0036)	0.4404 (0.0014)	0.4922 (0.0057)	0.4066 (0.0029)	0.6257 (0.0036)	0.4248 (0.0016)
	0.02	0.3121 (0.0041)	0.4789 (0.0016)	0.4078 (0.0042)	0.4388 (0.0014)	0.4751 (0.0061)	0.4025 (0.0031)	0.6118 (0.0043)	0.4203 (0.0016)
	0.01	0.3137 (0.0058)	0.4776 (0.0016)	0.3983 (0.0061)	0.4372 (0.0013)	0.4310 (0.0076)	0.4007 (0.0034)	0.5839 (0.0064)	0.4185 (0.0015)
(III)	0.03	0.2418 (0.0029)	0.4162 (0.0013)	0.3897 (0.0033)	0.4500 (0.0015)	0.5391 (0.0043)	0.4950 (0.0017)	0.6859 (0.0034)	0.5424 (0.0018)
	0.02	0.2535 (0.0037)	0.4155 (0.0012)	0.3913 (0.0042)	0.4502 (0.0015)	0.5222 (0.0048)	0.4958 (0.0016)	0.6758 (0.0041)	0.5401 (0.0018)
	0.01	0.2551 (0.0054)	0.4158 (0.0012)	0.3980 (0.0057)	0.4473 (0.0014)	0.4872 (0.0071)	0.4956 (0.0015)	0.6676 (0.0062)	0.5407 (0.0017)
(IV)	0.03	0.2983 (0.0032)	0.4479 (0.0013)	0.4599 (0.0034)	0.4761 (0.0015)	0.6700 (0.0035)	0.5089 (0.0016)	0.7569 (0.0030)	0.5267 (0.0017)
	0.02	0.3011 (0.0040)	0.4475 (0.0013)	0.4618 (0.0043)	0.4744 (0.0015)	0.6629 (0.0042)	0.5088 (0.0015)	0.7533 (0.0037)	0.5260 (0.0017)
	0.01	0.3657 (0.0056)	0.4783 (0.0014)	0.5470 (0.0064)	0.4970 (0.0016)	0.7633 (0.0052)	0.5124 (0.0015)	0.7857 (0.0048)	0.5167 (0.0016)
(V)	0.03	0.3668 (0.0035)	0.4801 (0.0015)	0.5432 (0.0036)	0.5015 (0.0017)	0.7762 (0.0033)	0.5178 (0.0016)	0.8027 (0.0029)	0.5229 (0.0017)
	0.02	0.3695 (0.0042)	0.4790 (0.0015)	0.5441 (0.0042)	0.5005 (0.0016)	0.7713 (0.0036)	0.5157 (0.0016)	0.7965 (0.0033)	0.5203 (0.0017)
	0.01	0.3657 (0.0056)	0.4783 (0.0014)	0.5470 (0.0064)	0.4970 (0.0016)	0.7633 (0.0052)	0.5124 (0.0015)	0.7857 (0.0048)	0.5167 (0.0016)
(VI)	0.03	0.2814 (0.0030)	0.4522 (0.0019)	0.3887 (0.0033)	0.4233 (0.0016)	0.5187 (0.0036)	0.4023 (0.0015)	0.5950 (0.0034)	0.4023 (0.0016)
	0.02	0.2773 (0.0037)	0.4516 (0.0019)	0.3861 (0.0039)	0.4221 (0.0015)	0.5055 (0.0045)	0.3990 (0.0015)	0.5828 (0.0041)	0.3981 (0.0016)
	0.01	0.2799 (0.0052)	0.4497 (0.0017)	0.3892 (0.0054)	0.4204 (0.0015)	0.4794 (0.0061)	0.3975 (0.0013)	0.5505 (0.0059)	0.3946 (0.0015)
(VII)	0.03	0.3022 (0.0031)	0.4520 (0.0014)	0.4640 (0.0035)	0.4872 (0.0015)	0.6852 (0.0038)	0.5260 (0.0017)	0.7706 (0.0032)	0.5443 (0.0019)
	0.02	0.3037 (0.0038)	0.4513 (0.0013)	0.4655 (0.0042)	0.4853 (0.0014)	0.6725 (0.0045)	0.5212 (0.0016)	0.7671 (0.0038)	0.5419 (0.0018)
	0.01	0.3039 (0.0052)	0.4507 (0.0013)	0.4677 (0.0059)	0.4828 (0.0014)	0.6513 (0.0060)	0.5187 (0.0017)	0.7516 (0.0052)	0.5388 (0.0018)
(VIII)	0.03	0.2599 (0.0030)	0.4513 (0.0014)	0.4253 (0.0035)	0.4877 (0.0016)	0.6358 (0.0038)	0.5240 (0.0017)	0.7363 (0.0033)	0.5451 (0.0019)
	0.02	0.2568 (0.0035)	0.4509 (0.0014)	0.4262 (0.0039)	0.4854 (0.0014)	0.6271 (0.0045)	0.5228 (0.0017)	0.7302 (0.0039)	0.5419 (0.0019)
	0.01	0.2591 (0.0047)	0.4501 (0.0013)	0.4315 (0.0055)	0.4826 (0.0014)	0.5912 (0.0065)	0.5184 (0.0016)	0.7109 (0.0054)	0.5389 (0.0017)

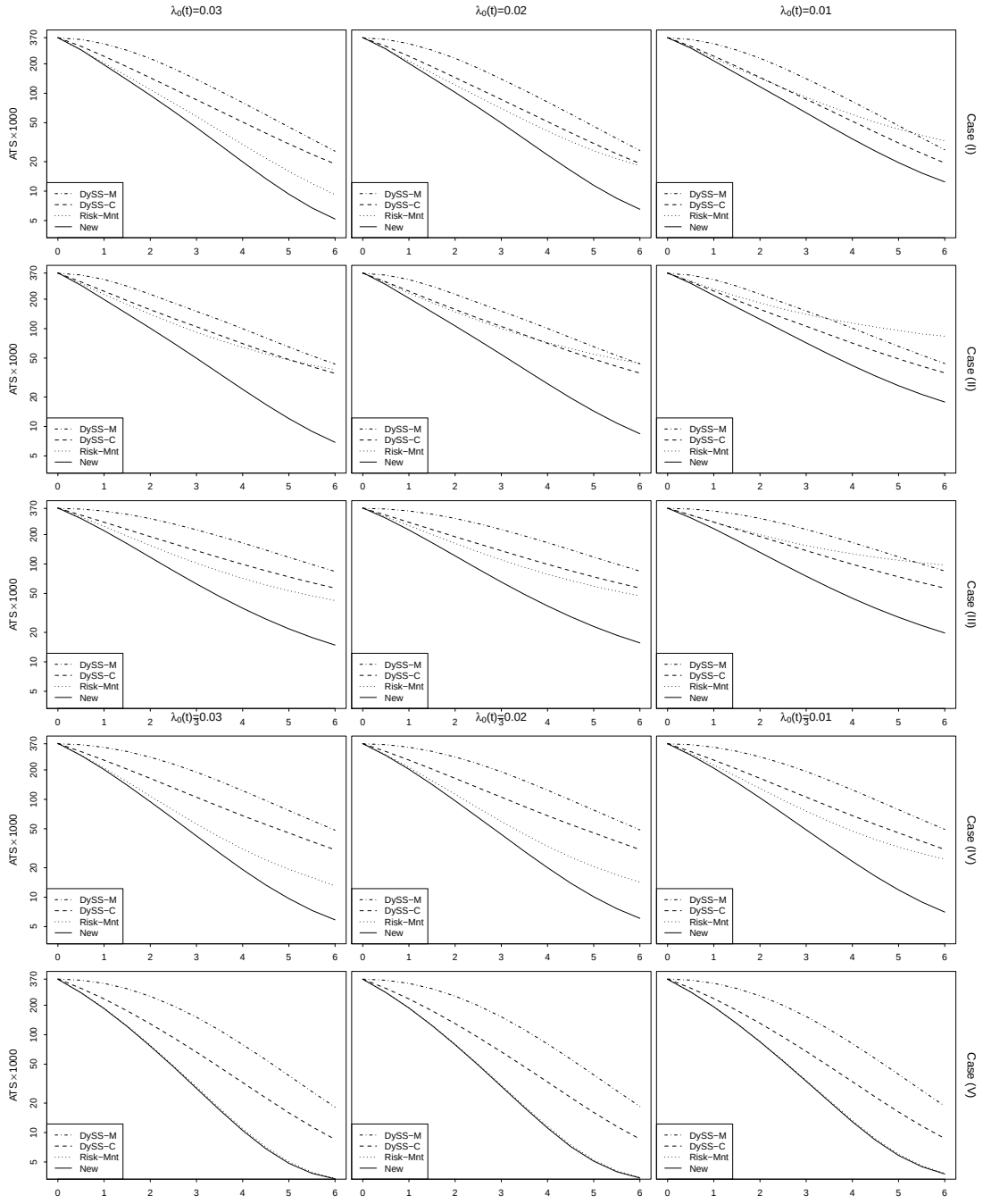

Figure 1: Calculated ATS_1 values of the four methods "DySS-M", "DySS-C", "Risk-Mnt", and "New" in cases (I)–(V) when ATS_0 is set to be 370×0.001.

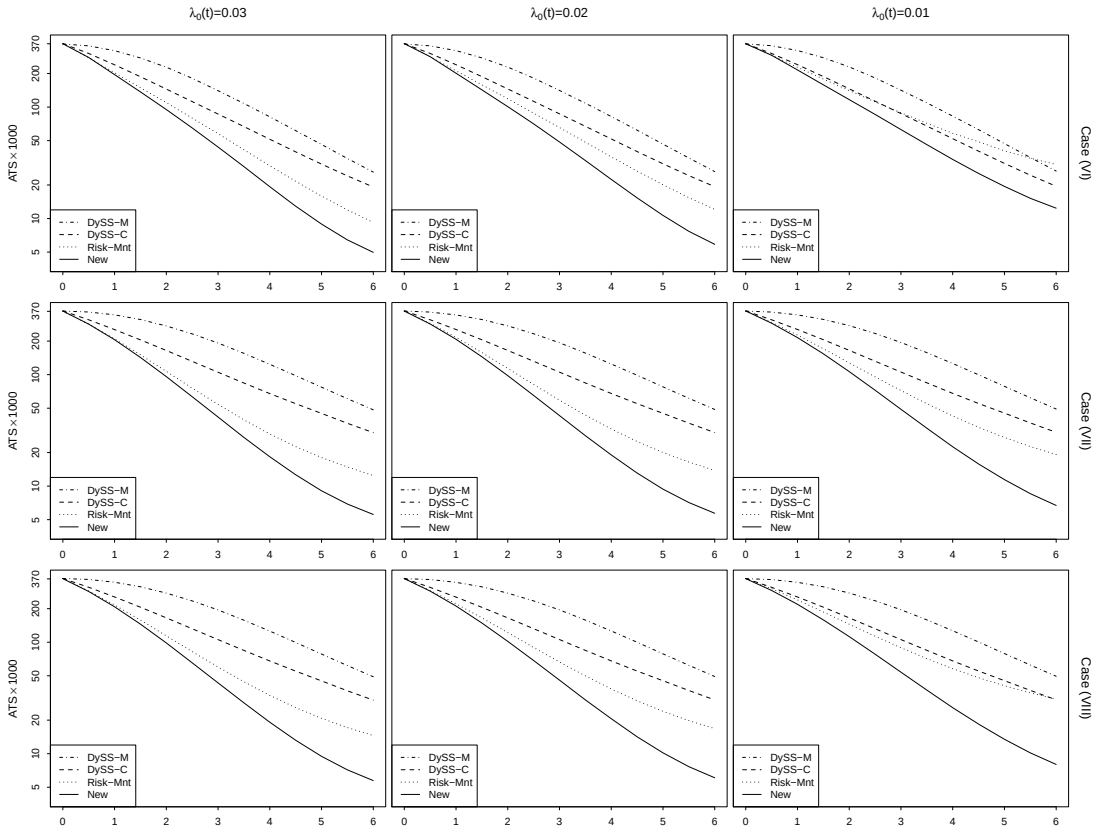

Figure 2: Calculated ATS_1 values of the four methods "DySS-M", "DySS-C", "Risk-Mnt", and "New" in cases (VI)–(VIII) when ATS_0 is set to be 370×0.001.

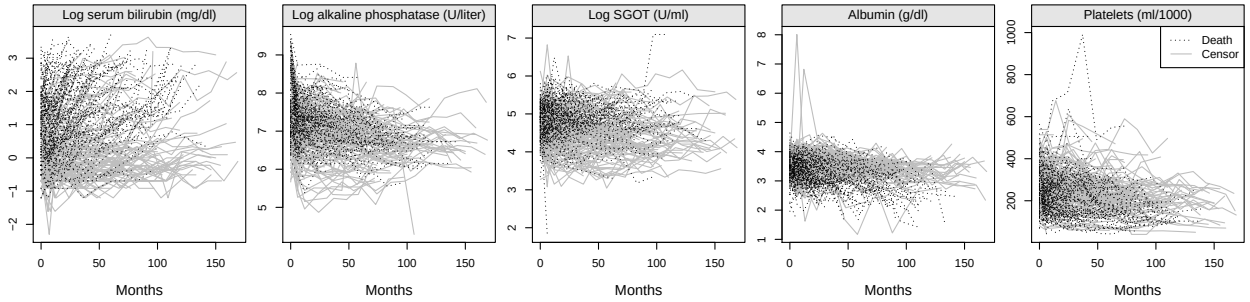

Figure 3: Five biomarkers in the PBC dataset. Dark dashed lines represent data of died patients and gray solid lines represent data of censored patients.

These two groups of patients are called "died patients" and "censored patients," respectively, in this section. Several biomarkers related to the disease progression were measured sequentially at the baseline time and the follow-up visits that were scheduled every 6 months. These biomarkers include log serum bilirubin (mg/dl), log alkaline phosphatase (U/liter), log serum glutamic-oxaloacetic transaminase (SGOT) (U/ml), albumin (g/dl), and platelets per cubic (ml/1000). The entire dataset is shown in Figure 3. On average, each patient made 6.23 clinical visits, resulting in a total of 1,945 observations for each biomarker. To apply the proposed method "New", the original dataset is randomly divided into two parts with equal numbers of died patients. The first part is used as a training dataset to estimate the joint model as discussed in Section 2, and the second part is used as a test dataset for evaluating the proposed method.

Our proposed joint models (1) and (2) are then estimated from the training dataset, where observations of the five biomarkers are the longitudinal data and times from enrollment to deaths are the times to the event (i.e., death in this case). After estimating the joint models, a test to check the proportional hazards assumption is performed using the method by Park and Qiu (2014). This method extended the tests in Grønnesby and Borgan (1996) and May and Hosmer (1998) to the joint-modeling setup. By this model diagnosis method, all patients in the training dataset are first divided into 10 groups of similar sizes based on the ranks of the weighted baseline measures $\sum_{k=1}^{p} \widehat{\beta}_k y_{ik}(t_{ik1})$. Then, the integrated martingale residual of the ith patient is defined to be

$$r_i = \int_0^{T_i} \widehat{\lambda}_0(t) \exp\{\widehat{\boldsymbol{\beta}}^{\mathsf{T}} \widehat{\mathbf{m}}_i(t)\}\, dt.$$

Let H_s be the sum of r_i for patients in the sth group, for $s = 1, 2, \ldots, 10$. Then, the test statistic

Table 3: Numbers of signals, computed ATS_1 values, and the corresponding FPR and TPR values of the four methods "DySS-M", "DySS-C", "Risk-Mnt" and "New" when they are applied to the test PBC data. Numbers in parentheses are the corresponding standard errors.

		Censored Patients		Died Patients			
Method	ATS_1	No Signal	Signal	No Signal	Signal	FPR	TPR
DySS-M	17.16	32	54	12	58	0.628 (0.052)	0.829 (0.045)
DySS-C	21.76	42	44	19	51	0.512 (0.054)	0.729 (0.053)
Risk-Mnt	10.73	48	38	8	62	0.442 (0.054)	0.886 (0.038)
New	8.98	54	32	8	62	0.360 (0.052)	0.886 (0.038)

to assess the proportional hazard assumption is defined to be

$$T_{\text{PH}} = (H_1, \ldots, H_9)\boldsymbol{\Sigma}_H^{-1}(H_1, \ldots, H_9)^{\mathsf{T}},$$

where $\boldsymbol{\Sigma}_H$ is an estimate of the covariance matrix of $(H_1, \ldots, H_9)^{\mathsf{T}}$. According to Park and Qiu (2014), T_{PH} would follow an asymptotic χ^2-distribution with 9 degrees of freedom under the proportional hazards assumption and some other regularity conditions. In this example, $T_{\text{PH}} = 13.257$ and the corresponding p-value is 0.151. So, there is no significant evidence in the training dataset to reject the proportional hazards assumption.

We then apply the proposed monitoring scheme "New" and the three competing methods "DySS-M", "DySS-C", and "Risk-Mnt" to sequentially monitor patients in the test dataset. The nominal ATS_0 value, defined as the average time from the beginning of online monitoring to the signal time among censored patients, is fixed to be 20 months for each method. The results are summarized in Table 3. From the table, it can be seen that "New" gives 32 false signals to a total of 86 censored patients, and it gives 62 true signals to a total of 70 died patients. Its false positive rate (FPR), which is equivalent to 1.0 minus the specificity, is 0.360, and its true positive rate (TPR), which is equivalent to the sensitivity, is 0.886. For its true signals, the ATS_1 value is 8.98 months, which is calculated to be the average time from the death of a patient to the signal time among died patients. By comparing "New" with the three competing methods, it can be seen that its FPR value is the lowest among the four methods, its TPR is higher than the ones of "DySS-M" and "DySS-C" and tied with the one of "Risk-Mnt", and its ATS_1 value is the shortest.

7 Concluding Remarks and Discussions

In this paper, a new DySS method for disease monitoring and disease early detection is proposed. The proposed new method first provides a formula for quantifying the disease risk using the joint modeling approach for analyzing the longitudinal and survival data in the training dataset, and then detects the disease in concern for individual people by sequentially monitoring the quantifed disease risks by a control chart. It has been shown that this method is more effective than several representative existing DySS methods. A practical procedure for choosing the weighting parameter λ used in the proposed method is also developed, which is not provided in the existing DySS research. In this paper, we focus primarily on the cases when the disease predictors are continuous variables. In many applications, one may encounter situations when such disease predictors are a combination of categorical, binary, count, and continuous variables. Modeling such data is often more challenging, and it is left for future research. Besides the local polynomial mixed-effects model considered in the proposed method, there have been some alternative recent methods for model estimation of nonparametric longitudinal models. For instance, Zhao et al. (2020) suggested using the penalized likelihood to simplify the the computation in model estimation, and Mauff et al. (2020) proposed a corrected two-stage approach for reducing the computation involved in the joint modeling. It might be possible to combine these methods with the proposed method in an appropriate way to further reduce its computation, which will be investigated in our future research. In addition, the survival model used in the proposed method relies on the proportional hazards assumption. Although some simulation examples have justified that our method is quite robust in certain cases when this assumption is violated, the model can be made more flexible to better describe the relationship between disease hazard rate and disease predictors. In Section 4.3, we have considered one such extension using a time-varying coefficients model. In the literature, there are some existing discussions about the case when the true hazard rate function is a nonparametric function of disease predictors. For instance, Huang (1999) used a generalized additive model to approximate the hazard function, and Fan et al. (1997) considered a local likelihood estimation method to estimate the hazard as a nonparametric function of disease predictors. Future research can explore the possibility to incorporate these more flexible survival models into the proposed method.

A Appendix

A.1 Model Estimation by local EM algorithm

In this part, we will give the full details of applying the local EM algorithm to model estimation. We first need to derive a formula for calculating the expectations of the random effects $\mathbf{b}_i(t)$ conditional on the observed data. Let $\boldsymbol{\psi}(\mathbf{b}_i(t))$ be some function of $\mathbf{b}_i(t)$. For a fixed t, we denote the expectation of $\boldsymbol{\psi}(\mathbf{b}_i(t))$ conditional on $\mathcal{Y}_i$ by $\widetilde{\mathrm{E}}_{i,t}[\boldsymbol{\psi}(\mathbf{b}_i(t))]$, and it can be computed by

$$\widetilde{\mathrm{E}}_{i,t}[\boldsymbol{\psi}(\mathbf{b}_i(t))] = \int_{\mathbb{R}^{pr}} \boldsymbol{\psi}(\mathbf{b}_i(t)) f_t(\mathbf{b}_i(t)|\mathcal{Y}_i)\, d\mathbf{b}_i(t).$$

In the above expression, by the Bayes rule, we have

$$f_t(\mathbf{b}_i(t)|\mathcal{Y}_i) \propto f_t(\mathcal{Y}_i|\mathbf{b}_i(t)) f_t(\mathbf{b}_i(t)).$$

It can also be checked that

$$\begin{aligned}
\log f_t(\mathcal{Y}_i|\mathbf{b}_i(t)) + \log f_t(\mathbf{b}_i(t)) &= -\frac{1}{2}\sum_{k=1}^{p}\sum_{j=1}^{n_{ik}} \log\{2\pi\sigma_k^2(t)\} K_h(t_{ikj}-t) \\
&- \frac{1}{2}\sum_{k=1}^{p}\sum_{j=1}^{n_{ik}} \frac{\left[y_{ik}(t_{ikj}) - \mathbf{e}_k^{\mathsf{T}}\mathbf{X}(t_{ikj}-t)\mathbf{b}_i(t)\right]^2}{\sigma_k^2(t)} K_h(t_{ikj}-t) \\
&- \frac{1}{2}\log\det(2\pi\boldsymbol{\Sigma}_b(t)) - \frac{1}{2}[\mathbf{b}_i(t) - \mathbf{c}(t) - \Delta_i\mathbf{d}(t)]^{\mathsf{T}}\boldsymbol{\Sigma}_b(t)^{-1}[\mathbf{b}_i(t) - \mathbf{c}(t) - \Delta_i\mathbf{d}(t)].
\end{aligned} \tag{15}$$

Therefore, because the expression (5) has a quadratic form of $\mathbf{b}_i(t)$, $\mathbf{b}_i(t)|\mathcal{Y}_i$ follows the multivariate normal distribution $N(\mathbf{m}_{b,i}(t), \mathbf{V}_{b,i}(t))$ with

$$\mathbf{m}_{b,i}(t) = \mathbf{c}(t) + \mathbf{V}_{b,i}(t)\sum_{k=1}^{p}\sum_{j=1}^{n_{ik}} \mathbf{X}(t_{ikj}-t)^{\mathsf{T}}\mathbf{e}_k\left[y_{ik}(t_{ikj}) - \mathbf{e}_k^{\mathsf{T}}\mathbf{X}(t_{ikj}-t)[\mathbf{c}(t) + \Delta_i\mathbf{d}(t)]\right] K_h(t_{ikj}-t)/\sigma_k^2(t),$$

$$\mathbf{V}_{b,i}(t) = \left[\boldsymbol{\Sigma}_b(t)^{-1} + \sum_{k=1}^{p}\sum_{j=1}^{n_{ik}} \mathbf{X}(t_{ikj}-t)^{\mathsf{T}}\mathbf{e}_k\mathbf{e}_k^{\mathsf{T}}\mathbf{X}(t_{ikj}-t) K_h(t_{ikj}-t)/\sigma_k^2(t)\right]^{-1}.$$

Thus, the expected log likelihood conditional on all observed data $\{\mathcal{Y}_i : i = 1, \ldots, m\}$ is

$$
\begin{aligned}
&\sum_{i \in R(t)} \widetilde{\mathrm{E}}_{i,t}\big[\log f_t(\mathcal{Y}_i|\mathbf{b}_i(t)) + \log f_t(\mathbf{b}_i(t))\big] \\
&\quad = -\frac{1}{2}\sum_{i=1}^{m}\sum_{k=1}^{p}\sum_{j=1}^{n_{ik}} \log\{2\pi\sigma_k^2(t)\} K_h(t_{ikj} - t) I(T_i \geq t) \\
&\qquad - \frac{1}{2}\sum_{i=1}^{m}\sum_{k=1}^{p}\sum_{j=1}^{n_{ik}} \frac{\widetilde{\mathrm{E}}_{i,t}\big[(y_{ik}(t_{ikj}) - \mathbf{e}_k^{\mathsf{T}}\mathbf{X}(t_{ikj} - t)\mathbf{b}_i(t))^2\big]}{\sigma_k^2(t)} K_h(t_{ikj} - t) I(T_i \geq t) \\
&\qquad - \frac{1}{2}\sum_{i=1}^{m} \widetilde{\mathrm{E}}_{i,t}\Big\{\big[\mathbf{b}_i(t) - \mathbf{c}(t) - \Delta_i\mathbf{d}(t)\big]^{\mathsf{T}}\boldsymbol{\Sigma}_b(t)^{-1}\big[\mathbf{b}_i(t) - \mathbf{c}(t) - \Delta_i\mathbf{d}(t)\big]\Big\} I(T_i \geq t) \\
&\qquad - \frac{1}{2}\sum_{i=1}^{m} \log\det(2\pi\boldsymbol{\Sigma}_b(t)) I(T_i \geq t).
\end{aligned}
$$

To maximize the above quantity, we can set its partial derivatives with respect to the time-varying parameters $\mathbf{c}(t)$, $\boldsymbol{\Sigma}_b(t)$ and $\sigma_k^2(t)$ to be zero, which results in the following parameter updating procedure in the maximization step:

$$
\mathbf{c}(t) \leftarrow \frac{\sum_{i:\Delta_i=0} \widetilde{\mathrm{E}}_{i,t}[\mathbf{b}_i(t)] I(T_i \geq t)}{\sum_{i:\Delta_i=0} I(T_i \geq t)} \tag{16}
$$

$$
\mathbf{d}(t) \leftarrow \frac{\sum_{i:\Delta_i=1} \widetilde{\mathrm{E}}_{i,t}[\mathbf{b}_i(t)] I(T_i \geq t)}{\sum_{i:\Delta_i=1} I(T_i \geq t)} - \mathbf{c}(t) \tag{17}
$$

$$
\boldsymbol{\Sigma}_b(t) \leftarrow \frac{\sum_{i=1}^{m} \widetilde{\mathrm{E}}_{i,t}\big[(\mathbf{b}_i(t) - \mathbf{c}(t) - \Delta_i\mathbf{d}(t))^{\otimes 2}\big] I(T_i \geq t)}{\sum_{i=1}^{m} I(T_i \geq t)} \tag{18}
$$

$$
\sigma_k^2(t) \leftarrow \frac{\sum_{i=1}^{m}\sum_{j=1}^{n_{ik}} \widetilde{\mathrm{E}}_{i,t}\big[(y_{ik}(t_{ikj}) - \mathbf{e}_k^{\mathsf{T}}\mathbf{X}(t_{ikj} - t)\mathbf{b}_i(t))^2\big] K_h(t_{ikj} - t) I(T_i \geq t)}{\sum_{i=1}^{m}\sum_{j=1}^{n_{ik}} K_h(t_{ikj} - t) I(T_i \geq t)}, \tag{19}
$$

where $\mathbf{x}^{\otimes 2} = \mathbf{x}\mathbf{x}^{\mathsf{T}}$ denotes the outer product of a vector with itself. From the distribution of $\mathbf{b}_i|\mathcal{Y}_i$, the quantities $\widetilde{\mathrm{E}}_{i,t}[\mathbf{b}_i(t)]$, $\widetilde{\mathrm{E}}_{i,t}\big[(\mathbf{b}_i(t) - \mathbf{c}(t) - \Delta_i\mathbf{d}(t))^{\otimes 2}\big]$ and $\widetilde{\mathrm{E}}_{i,t}\big[(y_{ik}(t_{ikj}) - \mathbf{e}_k^{\mathsf{T}}\mathbf{X}(t_{ikj} - t)\mathbf{b}_i(t))^2\big]$ in (16)–(19) have the expressions

$$
\begin{aligned}
\widetilde{\mathrm{E}}_{i,t}[\mathbf{b}_i(t)] &= \mathbf{m}_{b,i}(t), \\
\widetilde{\mathrm{E}}_{i,t}\big[(\mathbf{b}_i(t) - \mathbf{c}(t) - \Delta_i\mathbf{d}(t))^{\otimes 2}\big] &= \mathbf{V}_{b,i}(t), \\
\widetilde{\mathrm{E}}_{i,t}\big[(y_{ik}(t_{ikj}) - \mathbf{e}_k^{\mathsf{T}}\mathbf{X}(t_{ikj} - t)\mathbf{b}_i(t))^2\big] &= \mathbf{e}_k^{\mathsf{T}}\mathbf{X}(t_{ikj} - t)\mathbf{V}_{b,i}(t)\mathbf{X}(t_{ikj} - t)^{\mathsf{T}}\mathbf{e}_k \\
&\quad + \big[y_{ik}(t_{ikj} - t) - \mathbf{e}_k\mathbf{X}(t_{ikj} - t)\mathbf{m}_{b,i}(t)\big]^2.
\end{aligned}
$$

For each $t \in [0, \mathcal{T}]$, we can update the estimates of the time-varying parameters $\mathbf{c}(t)$, $\mathbf{d}(t)$, $\boldsymbol{\Sigma}_b(t)$ and $\sigma_k^2(t)$ by (16)–(19) until a convergence criterion holds.

References

Altman, N. S. (1990), "Kernel smoothing of data with correlated errors", *Journal of the American Statistical Association* **85,** 749–759.

Brown, E. R., Ibrahim, J. G., and DeGruttola, V. (2005), "A flexible B-spline model for multiple longitudinal biomarkers and survival", *Biometrics* **61,** 64–73.

Brumback, B. A., and Rice, J. A. (2005), "Smoothing spline models for the analysis of nested and crossed samples of curves", *Journal of the American Statistical Association* **93,** 961–976.

Chi, Y.-Y. and Ibrahim, J. G. (2006), "Joint models for multivariate longitudinal and multivariate survival data", *Biometrics* **62,** 432–445.

Epanechnikov, V. A. (1969), "Non-parametric estimation of a multivariate probability density," *Theory of Probability and Its Applications* **14,** 153–158.

Fan, J., Gijbels, I., and King, M. (1997), "Local likelihood and local partial likelihood in hazard regression," *Annals of Statistics* **25,** 1661–1690.

Grønnesby, J. K. and Borgan, Ø. (1996), "A method for checking regression models in survival analysis based on the risk score", *Lifetime Data Analysis* **2,** 315–328.

Huang, J. (1999), "Efficient estimation of the partly linear additive Cox model", *The Annals of Statistics*, **27,** 1536–1563.

Li, J. and Qiu, P. (2016), "Nonparametric dynamic screening system for monitoring correlated longitudinal data", *IIE Transactions*, **48**, 772–786.

Li, J. and Qiu, P. (2017), "Construction of an efficient multivariate dynamic screening system", *Quality and Reliability Engineering International*, **33**, 1969–1981.

Mauff, K., Erler, N. S., Kardys, I., and Rizopoulos, D. (2021), "Pairwise estimation of multivariate longitudinal outcomes in a Bayesian setting with extensions to the joint model", *Statistical Modelling*, **21**, 115–136.

May, S. and Hosmer, D. W. (1998), "A simplified method of calculating an overall goodness-of-fit test for the Cox proportional hazards model", *Lifetime Data Analysis* **4,** 109–120.

Park, K. Y. and Qiu, P. (2014), "Model selection and diagnostics for joint modeling of survival and longitudinal data with crossing hazard rate functions", *Statistics in Medicine* **33,** 4532–4546.

Qiu, P. and Xiang, D. (2014), "Univariate dynamic screening system: an approach for identifying individuals with irregular longitudinal behavior", *Technometrics*, **56**, 248–260.

Qiu, P. and Xiang, D. (2015), "Surveillance of cardiovascular diseases using a multivariate dynamic screening system", *Statistics in Medicine*, **34**, 2204–2221.

Qiu, P., Zi, X. and Zou, C. (2018), "Nonparametric dynamic curve monitoring," *Technometrics*, **60**, 386–397.

Rizopoulos, D. and Ghosh, P. (2011), "A Bayesian semiparametric multivariate joint model for multiple longitudinal outcomes and a time-to-event", *Statistics in Medicine* **30,** 1366–1380.

Shi, M., Weiss, R. E., and Taylor, J. M. (1996), "An analysis of paediatric CD4 counts for acquired immune deficiency syndrome using flexible random curves", *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **45,** 151–163.

Tian, L., Zucker, D., and Wei, L. J. (2005), "On the Cox model with time-varying regression coefficients", *Journal of the American statistical Association* **100,** 172–183.

Wu, H. and Zhang, J.-T. (2002), "Local polynomial mixed-effects models for longitudinal data", *Journal of the American Statistical Association* **97,** 883–897.

Wulfsohn, M. S. and Tsiatis, A. A. (1997), "A joint model for survival and longitudinal data measured with error", *Biometrics* **53,** 330–339.

Xiang, D., Qiu, P., and Pu, X. (2016), "Nonparametric regression analysis of multivariate longitudinal data", *Statistica Sinica* **23,** 769–789.

You, L. and Qiu, P. (2019), "Fast computing for dynamic screening systems when analyzing correlated data", *Journal of Statistical Computation and Simulation*, **89**, 379–394.

You, L. and Qiu, P. (2020), "An effective method for online disease risk monitoring", *Technometrics*, **62**, 249–264.

You, L. and Qiu, P. (2021), "Joint modeling of multivariate nonparametric longitudinal data and survival data: A local smoothing approach", Statistics in Medicine, **40,** 6689–6706.

Zhao, L., Chen, T., Novitsky, V., and Wang, R. (2020), "Joint penalized spline modeling of multivariate longitudinal data, with application to HIV-1 RNA load levels and CD4 cell counts", *Biometrics*, (in press).