Dynamic predictions using flexible joint models of longitudinal and time-to-event data

Jessica Barrett[a,‡] and Li Su[b,*,†,‡]

Joint models for longitudinal and time-to-event data are particularly relevant to many clinical studies where longitudinal biomarkers could be highly associated with a time-to-event outcome. A cutting-edge research direction in this area is dynamic predictions of patient prognosis (e.g., survival probabilities) given all available biomarker information, recently boosted by the stratified/personalized medicine initiative. As these dynamic predictions are individualized, flexible models are desirable in order to appropriately characterize each individual longitudinal trajectory. In this paper, we propose a new joint model using individual-level penalized splines (P-splines) to flexibly characterize the coevolution of the longitudinal and time-to-event processes. An important feature of our approach is that dynamic predictions of the survival probabilities are straightforward as the posterior distribution of the random P-spline coefficients given the observed data is a multivariate skew-normal distribution. The proposed methods are illustrated with data from the HIV Epidemiology Research Study. Our simulation results demonstrate that our model has better dynamic prediction performance than other existing approaches. © 2017 The Authors. *Statistics in Medicine* Published by John Wiley & Sons Ltd.

Keywords: P-splines; random effects; shared parameter models; survival analysis

1. Introduction

In many clinical trials and observational studies, longitudinal biomarker information is often collected together with information on a time-to-event outcome (e.g., patient survival). Joint modeling is becoming increasingly popular in characterizing the coevolution of the longitudinal and time-to-event processes [1]. Recently boosted by the stratified/personalized medicine initiative, a cutting-edge research direction in the joint modeling area is individualized dynamic predictions of patient prognosis (e.g., survival probabilities) using all available biomarker information. Pioneer works include Yu *et al*. [2], Proust-Lima and Taylor [3], Rizopoulos [4], and Taylor *et al*. [5].

As dynamic predictions for patient prognosis are individualized, flexible joint models are desirable in order to appropriately characterize the longitudinal process for each individual. In this paper, we propose a new flexible joint model with individual-level penalized splines (P-splines) [6] to characterize the coevolution of the longitudinal and time-to-event processes. One important strength of our model is that predicting survival probabilities becomes straightforward because the posterior distribution of individual-level (random) P-spline coefficients is a multivariate skew-normal distribution. We will start by reviewing relevant literature on flexible joint models, and then, we will describe the main idea of our approach, followed by details of our data example.

[a] *Strangeways Research Laboratory, Department of Public Health and Primary Care, University of Cambridge, Worts Causeway, Cambridge CB1 8RN, U.K.*

[b] *MRC Biostatistics Unit, School of Clinical Medicine, University of Cambridge Robinson Way, Cambridge CB2 0SR, U.K.*

[*] *Correspondence to: Li Su, MRC Biostatistics Unit, School of Clinical Medicine, University of Cambridge Robinson Way, Cambridge CB2 0SR, U.K.*

[†] *E-mail: li.su@mrc-bsu.cam.ac.uk*

[‡] *Both authors contributed equally and are joint first authors.*

1.1. Flexible joint models

In existing joint models of longitudinal and time-to-event (survival) data, it is often assumed that individual longitudinal trajectories are characterized by a linear model with random intercepts and random time slopes [7]. However, in long-term follow-up studies, individual longitudinal trajectories may not follow this simple linear model, which makes it challenging to examine the associated evolutions of the longitudinal and time-to-event processes. To overcome this problem, Brown *et al.* [8] proposed a flexible Bayesian B-spline model for the longitudinal process; Ding and Wang [9] also developed a nonparametric multiplicative random effects model to flexibly model the longitudinal process. More recently, Rizopoulos and Ghosh [10] proposed a Bayesian semiparametric multivariate joint model for multiple longitudinal outcomes and a time to event and discussed various parameterizations in the survival sub-model. Non-spline-based methods such as fractional polynomials have also been developed [11, 12], where the survival sub-model was based on cumulative hazard. None of the aforementioned works focused on dynamic predictions. Recently, Rizopoulos [7, Chapter 7] exemplified dynamic predictions based on a flexible joint model using B-splines with one interval knot in the longitudinal sub-model, but the smoothness (degree of freedom) of the B-splines was predefined and not chosen based on the data. Overall, frequentist approaches based on maximum likelihood estimation to flexible modeling of the longitudinal process in the joint modeling setting are still limited because of computational cost and complexity [13].

Recently, Barrett *et al.* [14] developed a joint model that allows more flexible random effect structures in the longitudinal sub-model. The key idea in their development was to discretize the time scale of the time-to-event outcome and let separate random effects within the discrete time intervals in the longitudinal sub-model (i.e., time-dependent random effects) be associated with the hazards of event occurrence in the corresponding time intervals. For example, a stationary Gaussian process was assumed for their random effect model. However, their specification of time-dependent random effects is somewhat limited by the fact that only a single random effect is used to characterize the evolution of the longitudinal process within a discrete time interval and the serial correlation between random effects is assumed to be stationary over time. Barrett *et al.* [14] did not investigate dynamic predictions based on their model.

1.2. Joint modeling with time-dependent random effects

In this paper, we propose a new flexible model for jointly modeling the longitudinal and time-to-event processes. The main idea is to use time-dependent random effects with non-stationary covariance structure, constructed by P-splines [15], to model time trends in each individual longitudinal trajectory. Specifically, building upon the model in [14], we use P-splines with a truncated linear basis [6] to model individual longitudinal trajectories, while population-level longitudinal trajectories can be modeled by P-splines with possibly different bases and knots. The smoothing parameters of both population-level and individual-level P-splines are chosen from the data in order to avoid over-fitting. Knots for the individual-level P-splines are located at the interval boundaries of the discretized time scale in order for the association between the longitudinal and survival sub-models to be easily interpreted. In this case, the individual-level P-spline coefficients act as shared parameters and are used to construct time-dependent random effects that represent the deviations of the intercepts and slopes of the individual longitudinal trajectories from the population-level trajectory within the corresponding discrete time interval, which is different from the single random effect setup used in [14]. Moreover, the covariance structure for these time-dependent random effects are non-stationary over time, unlike the stationary covariance structure used in [14].

1.3. Dynamic predictions

Model flexibility is even more important when performing individualized dynamic predictions. In practice, a simple linear model may fail to capture the nonlinear patterns in individual longitudinal trajectories for some patients, even if this model applies to the population-level longitudinal trajectories and the majority of individual longitudinal trajectories. Because dynamic predictions are individualized and need to appropriately account for possible nonlinearity in each individual longitudinal trajectory, flexible joint modeling approaches are certainly desirable. However, too much flexibility can also be harmful because of the danger of extrapolation [16]. In Section 5, we will use simulations to compare the dynamic prediction performance of our model with two other joint models with different degrees of flexibility.

A notable feature of our approach is that individualized dynamic predictions for survival probabilities over time are straightforward because under the proposed model the posterior distribution of the random
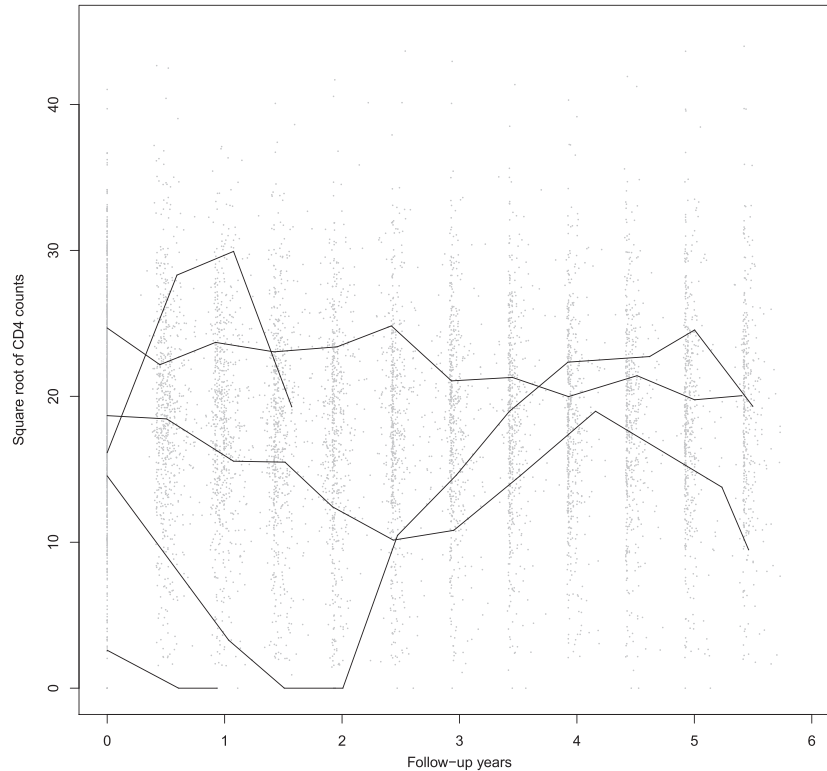
Figure 1. Observed (square root) longitudinal CD4 count data from the HIV Epidemiology Research Study with profiles from five selected participants highlighted.

P-spline coefficients given the observed data is a multivariate skew-normal distribution. This facilitates individualized predictions of survival probabilities [4] because no approximation (e.g., using Metropolis–Hastings algorithm) is needed to sample from this posterior distribution.

1.4. Example

This work is exemplified by data from the HIV Epidemiology Research Study (HERS) [17], with the aim of predicting HIV-related survival probabilities over time by jointly modeling longitudinal CD4 count process that reflected HIV disease progression in the study population. The HERS was a longitudinal study of 1310 women with, or at high risk for, HIV infection from 1993 to 2000. During the study, 12 visits were scheduled, where a variety of clinical, behavioral, and sociological outcomes were recorded approximately every 6 months. We will focus on the 850 women who were HIV positive at enrollment, and Figure 1 plots their observed CD4 count data over time (a square root transformation is used to reduce the right skewness in these data). It is clear that some individual longitudinal CD4 count trajectories had strong nonlinear patterns, which might be explained by the fact that the highly active antiretroviral therapy (HAART) was first introduced to the HERS population around 1997 and changed the disease progression dramatically. This phenomenon therefore motivates us to build a more flexible joint model to characterize individual longitudinal trajectories in order to improve predictions of HIV-related survival probabilities.

The rest of the paper is organized as follows. In Section 2, we introduce the proposed joint model. Estimation is described in Section 3, including the procedure for dynamic predictions. In Section 4, we apply the proposed methods to the HERS data. Simulation results are summarized in Section 5, and we conclude with a discussion in Section 6.

2. Joint model

Suppose that N independent patients are to be followed up over time. For the ith ($i = 1, \ldots, N$) patient, there is a longitudinal outcome process $\{Y_i(t)\}$, where $t \in \mathcal{T} = [0, T]$ is the time since enrollment and the constant T is determined by the potential maximum follow-up time where a longitudinal measurement can be taken in the study. Correspondingly, there is a p-dimensional covariate process $\{\mathbf{x}_i(t)\}$ associated with $\{Y_i(t)\}$. We assume $\{Y_i(t)\}$ is a continuous-time Gaussian process with a mean function $\mu_i(t)$ that is

dependent on $\mathbf{x}_i(t)$ and a variance–covariance function $\mathrm{cov}\{Y_i(t), Y_i(t') \mid \mathbf{x}_i(t), \mathbf{x}_i(t')\} = V_i(t, t')$ $(t \leqslant t')$. Parametric forms can be used for $V_i(t, t')$, for example, $V_i(t, t') = \sigma_\epsilon^2 \mathrm{I}(t = t')$ with $\mathrm{I}(\cdot)$ as an indicator function.

At the same time, a time-to-event outcome S_i is being observed and the occurrence of the event terminates the observation of $\{Y_i(t)\}$. Instead of using the continuous time scale $\mathcal{T}$ from the longitudinal process, we assume a discrete time scale $\mathcal{S} = \{1, 2, \ldots, M\}$ for this time-to-event outcome. However, it is assumed that there is a surjection $s(t)$ from $\mathcal{T}$ to $\mathcal{S}$, for example, $\mathcal{S}$ might be a partition of $\mathcal{T}$. Then, $\mathcal{S}$ is considered to be a series of time intervals. In the HERS example presented in Section 4, we partition the longitudinal measurement time by 6-month intervals and determine the occurrence of HIV-related deaths in these intervals. Further, let C_i be the potential censoring time for the ith patient. The observed event time is $S_i^* = \min(S_i, C_i)$, and the indicator for event occurrence is $\delta_i = \mathrm{I}(S_i \leqslant C_i)$. At continuous time points $t_{i1}, \ldots, t_{in_i}$ $(t_{in_i} \leqslant S_i^*)$, we also observe the longitudinal measurements $\mathbf{Y}_i = \{Y_i(t_{i1}), \ldots, Y_i(t_{in_i})\}^{\mathrm{T}}$.

2.1. Longitudinal sub-model

We assume the following model for the mean function $\mu_i(t)$ of $\{Y_i(t)\}$:

$$\mu_i(t) = \mathbf{x}_i(t)^{\mathrm{T}}\boldsymbol{\alpha} + m_i(t), \tag{1}$$

where $\boldsymbol{\alpha}$ is a $p \times 1$ vector of regression coefficients associated with covariates $\mathbf{x}_i(t)$ and $m_i(t)$ is the true underlying time trajectory for the ith patient after adjusting for $\mathbf{x}_i(t)$.

Following [6], $m_i(t)$ can be modeled by P-splines as follows:

$$m_i(t) = \sum_{l=0}^{M}(\beta_l + b_{il})B_l(t), \tag{2}$$

where $\{B_l(t)\} = \{1, t, (t - k_1)_+, \ldots, (t - k_{M-1})_+\}$ is the truncated linear basis with internal knots at $k_1, \ldots, k_{M-1}$ on $[0, T]$ and $(a)_+ = a \cdot \mathrm{I}(a \geqslant 0)$. The location of the knots $k_1, \ldots, k_{M-1}$ is determined by the boundary points of $\mathcal{S}$, the time intervals defined in the survival sub-model; thus, the number of internal knots is $M-1$. $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_M)^{\mathrm{T}}$ is the vector of P-spline coefficients that characterize the population-level longitudinal trajectory; $\mathbf{b}_i = (b_{i0}, b_{i1}, \ldots, b_{iM})^{\mathrm{T}}$ is the vector of random P-spline coefficients that determine the deviations of individual longitudinal trajectories from the population-level trajectory.

Note that we choose to use the truncated linear basis $\{B_l(t)\}$ because the time-dependent random effects constructed by P-spline coefficients and a truncated linear basis are easy to interpret when incorporated in the survival sub-model. For instance, in the HERS example in Section 4, we obtain the random intercept $W_{r0}(\mathbf{b}_i)$ at the rth interval $(r = 1, \ldots, M)$ as

$$W_{r0}(\mathbf{b}_i) = \begin{cases} b_{i0} & r = 1 \\ b_{i0} + b_{i1}k_1 & r = 2 \\ b_{i0} + b_{i1}k_{r-1} + \cdots + \sum_{l=2}^{r-1}b_{i,l}(k_{r-1} - k_{l-1}) & r = 3, \ldots, M \end{cases} \tag{3}$$

and the random slope $W_{r1}(\mathbf{b}_i)$ at the rth interval as

$$W_{r1}(\mathbf{b}_i) = b_{i1} + \cdots + b_{ir}; \tag{4}$$

see Figure S1 in the Supporting Information for a graphical illustration. In a similar manner, we have the population-level intercept and slope at the rth interval $W_{r0}(\boldsymbol{\beta})$ and $W_{r1}(\boldsymbol{\beta})$; therefore, the intercept and slope of $m_i(t)$ at the rth interval are $m_i(k_{r-1}) = W_{r0}(\boldsymbol{\beta}) + W_{r0}(\mathbf{b}_i)$ and $m_i'(k_{r-1}) = W_{r1}(\boldsymbol{\beta}) + W_{r1}(\mathbf{b}_i)$, respectively. Note that $k_0 = 0$ and $k_1, \ldots, k_{M-1}$ are the internal knots of the P-splines. In the corresponding survival sub-model, $m_i(k_{r-1})$ and $m_i'(k_{r-1})$ are included as time-varying covariates and their regression coefficients represent the associations of the level of CD4 count at the beginning of the time intervals and the progression rate of CD4 count within the time intervals with the corresponding conditional survival probabilities in these intervals.

In practice, a more complex basis such as B-splines or low-rank thin-plate splines (with better numerical properties) could also be used for the population-level trajectory [15]; the penalized likelihood estimation procedure in Section 3 can easily be adapted to use any basis for fitting the population-level

trajectory. For individual trajectories, because in most applications the number of data points for each patient is not large, the truncated linear basis is generally flexible enough to characterize the essential patterns of each individual longitudinal trajectory.

For both population and individual-level P-splines with truncated linear basis in our model, knot locations could be chosen to lie anywhere on the continuous time scale. In this paper, we choose to fix the knot locations at the discretization points in the survival sub-model because of the following reasons. First, in practice, a summary of the rate of change of the longitudinal outcome across a discrete time interval is needed to define an association of the longitudinal outcome with the survival process. The association between the longitudinal and survival processes is therefore easier to interpret if knots lie on discrete time interval boundaries because the time slope of an individual's longitudinal trajectory is then *constant* within the time intervals given our setup of P-splines with truncated linear basis. Second, using the sample quantiles of the longitudinal measurement times in the observed data to define the knot locations, as is common practice in semiparametric regression literature [15], could be problematic in our scenario because the knots would be closer together at earlier follow-up times because of selection bias by the survival outcome. Because the longitudinal sub-model is intended to characterize the longitudinal process if no truncation of the survival outcome occurs and the joint modeling approach is adopted to correct the selection bias due to the survival outcome, we choose the knot locations and discretization points according to the planned longitudinal measurement schedule, thereby avoiding dependence of the longitudinal sub-model specification on the observed survival outcomes.

2.2. Survival sub-model

Following [14], we assume a probit model for the discrete hazard rate of the event $\lambda_{ir} = P(S_i = r \mid S_i \geqslant r)$ in the rth interval ($r = 1, \ldots, M$),

$$\lambda_{ir} = 1 - \Phi\left(\tilde{\mathbf{x}}_{ir}^T \tilde{\boldsymbol{\alpha}} + \boldsymbol{\gamma}_r^T \mathbf{W}_{ir}\mathbf{b}_i\right), \tag{5}$$

where $\Phi(\cdot)$ is the standard normal cumulative distribution function, $\tilde{\mathbf{x}}_{ir}$ is a $\tilde{q} \times 1$ vector of covariates (possibly time varying) with regression coefficients $\tilde{\boldsymbol{\alpha}}$. $\mathbf{W}_{ir}\mathbf{b}_i$ is a $q \times 1$ vector of linear combinations of $\mathbf{b}_i$ (e.g., in the HERS example, we have $\mathbf{W}_{ir}\mathbf{b}_i = (m_i(k_{r-1}), m_i'(k_{r-1}))^T$ and $q = 2$) and $\boldsymbol{\gamma}_r$ is an association parameter vector that relates the longitudinal and time-to-event processes.

Depending on the applications, various parameterizations for $\mathbf{W}_{ir}\mathbf{b}_i$ can be used; see discussions in [7] and [10]. For example, we use $m_i(k_{r-1})$ and $m_i'(k_{r-1})$ in the HERS example in Section 4, as it is believed that the survival probabilities depend on the disease progression, but we only allow disease progression up to the end of the rth interval to be associated with the survival probability at the rth interval. Interactions between $m_i(k_{r-1})$, $m_i'(k_{r-1})$ and baseline covariates could also be included.

2.3. Random effects

Following the shared parameter modeling framework, we assume that the longitudinal process $\{Y_i(t)\}$ and the time-to-event outcome S_i are independent conditional on the P-spline coefficients $\mathbf{b}_i$ and all covariates. Further, we assume that $\mathbf{b}_i$ are also independent of all covariates and

$$\begin{bmatrix} b_{i0} \\ b_{i1} \\ b_{i2} \\ \vdots \\ b_{iM} \end{bmatrix} \sim N\left(\mathbf{0}, \boldsymbol{\Sigma} = \begin{bmatrix} \sigma_0^2 & \rho\sigma_0\sigma_1 & \mathbf{0} \\ \rho\sigma_0\sigma_1 & \sigma_1^2 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \sigma_2^2\mathbf{I}_{M-1} \end{bmatrix}\right), \tag{6}$$

where $\mathbf{I}_{M-1}$ is the $(M-1)$-dimensional identity matrix. Note that the P-spline coefficients $\mathbf{b}_i$ are used to construct time-dependent random effects such as $W_{r0}(\mathbf{b}_i)$ and $W_{r1}(\mathbf{b}_i)$ in (3) and (4). Therefore, assuming $(b_{i2}, \ldots, b_{iM}) \sim N(\mathbf{0}, \sigma_2^2\mathbf{I}_{M-1})$ does not mean that $(W_{10}(\mathbf{b}_i), \ldots, W_{M0}(\mathbf{b}_i))$ and $(W_{11}(\mathbf{b}_i), \ldots, W_{M1}(\mathbf{b}_i))$ are independent. In fact, from the functional form described in (3) and (4), it is apparent that these time-dependent random effects are correlated with each other over time. Unlike the stationary covariance structure specified in [14], the covariance structure for time-dependent random effects in our model is non-stationary over time; for example, it is easy to see that $\mathrm{cov}(W_{10}(\mathbf{b}_i), W_{20}(\mathbf{b}_i)) = \sigma_0^2 + k_1\rho\sigma_0\sigma_1$ but $\mathrm{cov}(W_{20}(\mathbf{b}_i), W_{30}(\mathbf{b}_i)) = \sigma_0^2 + k_1k_2\sigma_1^2 + (k_1 + k_2)\rho\sigma_0\sigma_1$. In sum, the serial correlation in the longitudinal

process is characterized by the time-dependent random effects with a non-stationary covariance structure over time.

Note that σ_2^2 is the smoothing parameter that penalizes the non-smoothness of the individual-level P-splines (i.e., large values of $b_{i2}, \ldots, b_{iM}$)) and will be estimated by maximizing the marginal likelihood [6]. The B-spline approach applied in existing joint models [7] used a small number of knots (say 1–3) for each individual longitudinal trajectory but did not adapt the smoothness of the B-splines to the data. Therefore, using high-degree polynomial terms (e.g., cubic terms) in B-splines could potentially lead to over-fitting. In Section 5, we will conduct a simulation study to specifically compare the dynamic prediction performance of the joint model based on our P-spline approach with a cubic spline approach.

The heterogeneity between individual longitudinal trajectories is determined by $\sigma_0, \sigma_1, \sigma_2$. In the HERS example in Section 4, we parameterize $\sigma_0, \sigma_1, \sigma_2$ using the log transformations and ρ using Fisher's z-transform, $z(x) = \frac{1+x}{1-x}$.

3. Estimation

The estimation for the joint model proposed in Section 2 is based on a maximum penalized likelihood approach that maximizes the likelihood function corresponding to the joint distribution of the longitudinal and time-to-event outcomes $\{\mathbf{Y}_i, S_i^* = s, \delta_i\}$ times a penalty term for smoothing the population-level P-spline coefficients $(\beta_2, \ldots, \beta_M)$.

3.1. Likelihood

Specifically, the likelihood contribution from the ith patient is

$$\mathcal{L}_i(\theta \mid \mathbf{Y}_i, S_i^* = s, \delta_i) = \int f(\mathbf{Y}_i \mid \mathbf{b}_i; \theta) f(s, \delta_i \mid \mathbf{b}_i; \theta) f(\mathbf{b}_i; \theta) d\mathbf{b}_i, \tag{7}$$

where θ denotes all unknown parameters. Let $\mathbf{x}_i = (\mathbf{x}(t_{i1}), \ldots, \mathbf{x}(t_{in_i}))^\mathrm{T}$ and $\mathbf{B}_i$ be the $n_i \times (M+1)$ matrix for the truncated linear basis $\{B_l(t)\} = \{1, t, (t-k_1)_+, \ldots, (t-k_{M-1})_+\}$ evaluated at time points $t_{i1}, \ldots, t_{in_i}$. The likelihood from the longitudinal part is

$$f(\mathbf{Y}_i \mid \mathbf{b}_i; \theta) = \exp\left\{-\log(2\pi)n_i/2 - \log(|\mathbf{V}_i|)/2 - (\mathbf{Y}_i - \mu_i)^\mathrm{T}\mathbf{V}_i^{-1}(\mathbf{Y}_i - \mu_i)/2\right\},$$

where $\mu_i = \mathbf{x}_i\alpha + \mathbf{B}_i\beta + \mathbf{B}_i\mathbf{b}_i$ and $\mathbf{V}_i$ is the covariance matrix obtained by evaluating the covariance function $V_i(t, t')$ at time points $t_{i1}, \ldots, t_{in_i}$. The likelihood of the survival part is

$$f(s, \delta_i \mid \mathbf{b}_i; \theta) = \left\{\prod_{r=1}^{s-1} \Phi\left(\tilde{\mathbf{x}}_{ir}^\mathrm{T}\tilde{\alpha} + \gamma_r^\mathrm{T}\mathbf{W}_{ir}\mathbf{b}_i\right)\right\} \left\{\Phi\left(\tilde{\mathbf{x}}_{is}^\mathrm{T}\tilde{\alpha} + \gamma_s^\mathrm{T}\mathbf{W}_{is}\mathbf{b}_i\right)\right\}^{1-\delta_i} \times \left\{1 - \Phi\left(\tilde{\mathbf{x}}_{is}^\mathrm{T}\tilde{\alpha} + \gamma_s^\mathrm{T}\mathbf{W}_{is}\mathbf{b}_i\right)\right\}^{\delta_i}. \tag{8}$$

The density $f(\mathbf{b}_i; \theta)$ is the multivariate normal $N(\mathbf{0}, \Sigma)$ defined in Section 2.3. Using the multivariate skew-normal results [14, 18], it can be shown that the likelihood in (7) can be written analytically. Details are given in the Supporting Information.

3.2. Maximum penalized likelihood estimation

Finally, we incorporate the penalty term for smoothing the population-level P-splines coefficients $\tilde{\beta} = (\beta_2, \ldots, \beta_M)$ to the likelihood function as follows:

$$\mathcal{PL} = \prod_{i=1}^{N} \mathcal{L}_i(\theta \mid \mathbf{Y}_i, S_i^* = s, \delta_i) \cdot \exp(-\lambda\tilde{\beta}^\mathrm{T}\tilde{\beta}),$$

where λ $(\lambda > 0)$ is the smoothing parameter [6].

For a fixed value of λ, estimation of θ can be performed by numerical maximization of the penalized likelihood. The variance–covariance matrix of the maximum penalized likelihood estimates $\hat{\theta}$ can be estimated by the inverse of the observed Fisher information matrix. To choose an optimal value of λ, we first calculate the degree of freedom of population-level P-splines for a fixed λ as $df(\lambda) = \text{tr}\{(\tilde{\mathbf{B}}^{\text{T}}\tilde{\mathbf{B}} + 2\lambda\hat{\sigma}_{\epsilon}^2\mathbf{D})^{-1}\tilde{\mathbf{B}}^{\text{T}}\tilde{\mathbf{B}}\}$, where $\tilde{\mathbf{B}} = (\mathbf{B}_1^{\text{T}}, \ldots, \mathbf{B}_N^{\text{T}})^{\text{T}}$, $\mathbf{D} = \begin{bmatrix} \mathbf{0}_{2\times 2} & \mathbf{0}_{2\times(M-1)} \\ \mathbf{0}_{(M-1)\times 2} & \mathbf{I}_{(M-1)\times(M-1)} \end{bmatrix}$, and $\hat{\sigma}_{\epsilon}^2$ is the maximum penalized likelihood estimate for the error variance σ_{ϵ}^2 assuming that $V_i(t, t') = \sigma_{\epsilon}^2 \text{I}(t = t')$ [19]. Then, the Akaike information criterion (AIC) [19] is

$$AIC(\lambda) = -2\log(\hat{\mathcal{L}}) + 2 \cdot df(\lambda) + 2 \cdot \dim(\tilde{\theta}),$$

where $\hat{\mathcal{L}} = \prod_{i=1}^{N} \mathcal{L}_i(\hat{\theta} \mid \mathbf{Y}_i, S_i^* = s, \delta_i)$ is the likelihood from the joint model evaluated at the maximum penalized likelihood estimates $\hat{\theta}$ and $\tilde{\theta}$ is a subset of θ by excluding $\tilde{\beta}$. We minimize $AIC(\lambda)$ to estimate the optimal λ.

3.3. Dynamic predictions of survival probabilities

In this section, we describe the procedure to perform dynamic predictions of survival probabilities based on the joint model in Section 2. Suppose that we are interested in predicting the conditional survival probability of surviving the sth interval given survival over the rth $(r < s)$ interval for a new patient i, $\pi_i(s \mid r) = \text{P}(S_i > s \mid S_i > r, \mathbf{Y}_i\{t(r)\}, \mathcal{D}_N; \theta, \lambda)$, where $\mathbf{Y}_i\{t(r)\}$ is the series of longitudinal measurements provided by this patient up to $t(r)$, the right end point of the r interval. $\mathcal{D}_N = \{\mathbf{Y}_i, S_i^* = s, \delta_i; i = 1, \ldots, N\}$ is the sample on which the joint model was fitted.

For a joint model with random intercept and slope, Rizopoulos [4] proposed two estimators of $\pi_i(s \mid r)$: (i) by evaluating $\pi_i(s \mid r)$ at $\hat{\theta}$ and $\hat{\mathbf{b}}_i$, the empirical Bayes estimates of $\mathbf{b}_i$ given $S_i > r, \mathbf{Y}_i\{t(r)\}$ and (ii) by sampling $\theta^{(l)}$ from its asymptotic distribution $N(\hat{\theta}, \text{vâr}(\hat{\theta}))$ and $\mathbf{b}_i^{(l)}$ from the posterior distribution $f(\mathbf{b}_i \mid S_i > r, \mathbf{Y}_i\{t(r)\}, \theta^{(l)})$ and then obtaining an estimate (e.g., median) from the samples of $\pi_i(s \mid r)$ evaluated at $\theta^{(l)}, \mathbf{b}_i^{(l)}$. The simulation results of [4] showed that the two estimators both performed well in terms of dynamic prediction accuracy.

In our model, the smoothing parameter λ is estimated separately from the other model parameters. Therefore, it seems problematic to sample from the asymptotic distribution $N(\hat{\theta}, \text{vâr}(\hat{\theta}))$ because the population-level P-spline coefficients $\tilde{\beta}$ (as a subset of θ) are also implicitly determined by λ. On the other hand, conditioning on $\hat{\theta}$ and λ, we can draw from $\mathbf{b}_i^{(l)} \sim f(\mathbf{b}_i \mid S_i > r, \mathbf{Y}_i\{t(r)\}, \hat{\theta}, \lambda)$ and calculate $\hat{\pi}_i(s \mid r) = \prod_{r*=r+1}^{s} \Phi\left(\tilde{\mathbf{x}}_{ir*}^{\text{T}}\hat{\alpha} + \gamma_{r*}^{\text{T}}\mathbf{W}_{ir*}\hat{\mathbf{b}}_i\right)$, where $\hat{\mathbf{b}}_i$ could be the sample mean, median, or mode of $\mathbf{b}_i^{(l)}$ $(l = 1, \ldots, L)$ and L is the number of Monte Carlo samples. This is the dynamic prediction procedure we will use in the analysis of the HERS data and in the simulations.

Note that $f(\mathbf{b}_i \mid S_i > r, \mathbf{Y}_i\{t(r)\}, \hat{\theta}, \lambda)$ can be shown to be the density of a multivariate skew-normal distribution (see details in the Supporting Information). This simplifies the prediction procedure as no approximation, for example, using a Metropolis–Hastings algorithm, is needed for sampling from this distribution [4].

4. Application to the HIV Epidemiology Research Study data

In this section, we apply the proposed methods to the HERS data introduced in Section 1. During the follow-up of the HERS, there were 106 HIV-related deaths, which censored the longitudinal CD4 processes for these patients. Censoring by dropout also occurred, which was possibly related to disease progression. Previous analyses of the HERS CD4 count data [20] did not distinguish censoring by death and dropout. In our analysis, we will assume that given the random effects that characterize the individual longitudinal CD4 count process, the dropout time and the HIV-related survival time were independent. In other words, we will focus on modeling HIV-related survival time and treat dropout as independent censoring conditional on random effects. For those women who actually finished 12 scheduled visits, the HIV-related survival times are treated as administratively censored.

The maximum follow-up time was 2093 days in the HERS data, and we partition the follow-up period into 12 intervals. Except for the first interval that is 3 months from enrollment, the remaining 11 intervals are equally spaced every 6 months, which ensures that each interval approximately contains

one scheduled CD4 count measurement. We also standardize the square root of CD4 count by taking $(\sqrt{CD4} - 18)/7$ to facilitate computation.

4.1. Models

We fit two joint models to these data. In both joint models, we assume that the covariance function of the longitudinal process is $V_i(t, t') = \sigma_\epsilon^2 I(t = t')$. We use $V_i(t, t') = \sigma_\epsilon^2 I(t = t')$ to account for measurement errors only because the specified random effects structure in (6) is used to capture the non-stationary serial correlation over time for longitudinal data. In practice, other parametric models can be used for $V_i(t, t')$, for instance, $V_i(t, t') = \sigma^2 \exp(-\phi|t' - t|)$, in order to characterize the remaining serial correlations.

In the first joint model ('Model 1'), we assume that the mean function in the longitudinal CD4 count process is

$$\mu_i(t) = \sum_{l=0}^{12} (\beta_l + b_{il}) B_l(t),$$

where t is the follow-up time in days (scaled by 2093 such that $t \in [0, 1]$), $\{B_l(t)\} = \{1, t, (t - k_1)_+, \ldots, (t - k_{11})_+\}$ is the truncated linear basis with knots corresponding to $3, 9, \ldots, 63$ months before scaling, and $\mathbf{b}_i = (b_{i0}, \ldots, b_{i,12})$ are the corresponding random P-spline coefficients that follow the distribution specified in (6). Knot locations were chosen based on the planned visit schedule.

For comparison purposes, in the second joint model ('Model 2'), we assume that the mean function of the longitudinal process is

$$\mu_i(t) = \beta_0 + \beta_1 t + b_{i0} + b_{i1} t,$$

and

$$\begin{bmatrix} b_{i0} \\ b_{i1} \end{bmatrix} \sim N \left(\mathbf{0}, \begin{bmatrix} \sigma_0^2 & \rho\sigma_0\sigma_1 \\ \rho\sigma_0\sigma_1 & \sigma_1^2 \end{bmatrix} \right)$$

are the random intercept and time slope for the ith patient.

In the survival sub-model of Model 1, based on some preliminary investigations and the findings in [17], we assume that

$$\lambda_{ir} = 1 - \Phi\{\alpha_0 + \alpha_1 \tilde{r} + \alpha_2 \tilde{r}^2 + \alpha_3 \text{age}_i + \alpha_4 V_{2i} + \alpha_5 V_{3i} + \alpha_6 V_{4i} + \gamma_0 m_i(k_{r-1}) + \gamma_1 m'_i(k_{r-1})\},$$

where $\tilde{r}$ is the time at the start of the rth time interval, age_i is the age at enrollment (standardized by taking $(\text{age}_i - 35)/7$) and V_{2i}, V_{3i}, V_{4i} are the indicator variables for HIV viral load groups $(500, 5000]$, $(5000, 30\,000]$, $(30\,000, \infty)$ at enrollment, respectively. The intercept $m_i(k_{r-1})$ and slope $m'_i(k_{r-1})$ of the current time interval are incorporated as time-varying covariates. For Model 2, we assume the same survival sub-model except that $m(k_{r-1}) = \beta_0 + b_{i0} + (\beta_1 + b_{i1})k_{r-1}$ and $m'(k_{r-1}) = \beta_1 + b_{i1}$ because the time slope is assumed constant.

We use the estimation methods in Section 3 to fit the two joint models and perform dynamic predictions based on these fitted models following the procedure described in Section 3.3.

4.2. Summary of fitted models

Figure 2 presents the estimated population-level longitudinal CD4 count trajectories from the two joint models. Table I summarizes the results of survival sub-models and variance components from Models 1 and 2. Overall, the values of AIC indicate that Model 1 provides much better fit to the observed data. The optimal value for the smoothing parameter λ is 0.4, which means that the effective number of parameters in the population-level P-splines is 9.4 (note that the number of population-level P-spline coefficients for penalization is 11).

The estimated population-level longitudinal CD4 count trajectory from Model 1 suggests that the disease progression in the HERS cohort (decline of CD4 count) was slowing down in the middle of follow-up when the HAART was introduced. In both models, after adjusting for enrollment age, viral load group, and the time at the start of the interval, the conditional probability of surviving each time interval was positively associated with the current intercept and time slope of the CD4 count. However, the estimates for γ_0 and γ_1 from Model 1 are both larger than those from Model 2, especially for γ_1 (0.277 vs. 0.172) with a difference of almost two standard errors. This suggests that estimates from Model 2 are possibly attenuated because of its less flexible modeling of individual longitudinal trajectories.
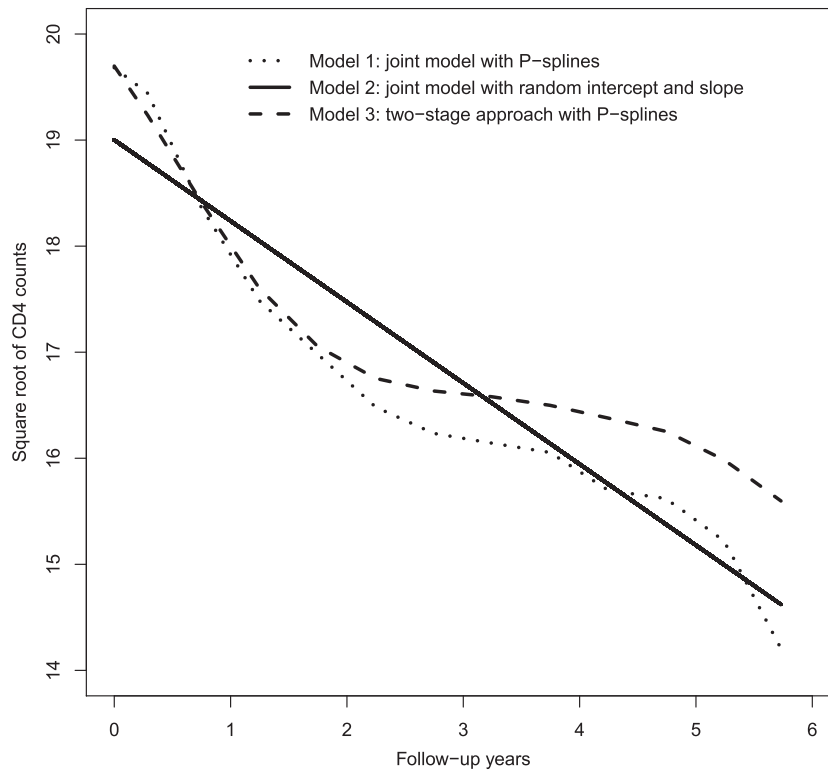
Figure 2. Estimated population-level CD4 count longitudinal trajectories from the two-stage approach and the two joint models fitted to the HIV Epidemiology Research Study data.

Table I. Parameter estimates, standard errors, and model comparison results from the two joint models fitted to the HERS data.

Parameter		Model 1	Model 2
Survival	α_0	4.006 (0.342)	3.573 (0.272)
	α_1	−2.328 (0.932)	−1.643 (0.660)
	α_2	2.701 (1.028)	1.995 (0.750)
	α_3	−0.211 (0.062)	−0.194 (0.056)
	α_4	−0.345 (0.242)	−0.349 (0.240)
	α_5	−0.606 (0.248)	−0.603 (0.247)
	α_6	−0.536 (0.250)	−0.569 (0.259)
	γ_0	0.775 (0.090)	0.751 (0.079)
	γ_1	0.277 (0.055)	0.172 (0.058)
Others	σ_ϵ	0.353 (0.004)	0.386 (0.004)
	σ_0	0.916 (0.024)	0.931 (0.025)
	σ_1	0.778 (0.081)	1.146 (0.043)
	σ_2	0.658 (0.038)	—
	ρ	0.123 (0.059)	−0.121 (0.046)
log likelihood		−5543.876	−5761.138
AIC		11 138.54	11 552.28
λ		0.4	—
df(λ)		9.4	—

HERS, HIV Epidemiology Research Study; AIC, Akaike information criterion.

In summary, the HIV-related survival in the HERS cohort was associated with younger age and lower viral load at enrollment as well as current higher level and increasing rate of CD4 count, which is consistent with the findings in [17].

4.3. Dynamic predictions

In this section, we demonstrate dynamic prediction based on our flexible joint model. In particular, we exemplify it by making predictions on the conditional probabilities of surviving the next one, two, and three intervals given all baseline information and available CD4 count measurements up to the cutoff time for predictions as well as the fact that a patient was still under follow-up at this cutoff time.

For comparison purpose, we also perform dynamic predictions using a survival model with the longitudinal outcome as a time-varying covariate as well as a two-stage approach. The survival model with the time-varying covariate will follow the same structure as in the joint models, but Y_{ir}, instead of estimates of $m_i(k_{r-1})$ and $m_i'(k_{r-1})$, is incorporated as a time-varying covariate. The dynamic prediction procedure for this approach is similar to those used for the joint models, except that the last observed outcome Y_{ir} (instead of random effect estimates) is used in the fitted survival model for prediction.

In the two-stage approach, first we fit a linear mixed model with P-splines to the observed longitudinal data using the same specification as for Model 1. The computation is carried out by the `lme` function in the R package `nlme`. Figure 2 also gives the estimated population-level longitudinal CD4 count trajectory from the two-stage approach, which overestimates the CD4 count level at later follow-up time because it ignores the selection through survival. Using the empirical Bayes estimates of the random effects from the fitted linear mixed model, we then fit a survival model with the same specification as in Models 1 and 2. Based on the parameter point estimates from the linear mixed model, we obtain the empirical Bayes estimates of the random effects for the patients we would like to make predictions for. Finally, using these random effect estimates and the fitted survival model, we produce predicted survival probabilities over time for these patients. Note that, unlike in the joint models, the posterior distribution of the random effects used to generate empirical Bayes estimates in the two-stage approach will not involve the observed survival data.

As an example, we consider patient 26 who was 37 years old with viral load (5000, 30 000] at enrollment. The left sides of the panels of Figure 3 present the observed (standardized) square root CD4 counts and estimated individual longitudinal trajectories for patient 26 up to the prediction time $r = 3, 5, 7, 9$, respectively. The right sides provide predicted conditional probabilities of HIV survival after the next one, two, and three time intervals at the prediction time $r = 3, 5, 7, 9$, respectively. For the joint models, we use medians of 200 samples from the posterior of $\mathbf{b}_i$. The same figure presented on the probit scale is given in Figure S2. At $r = 3$, the predicted conditional survival probabilities are similar for all models, while at $r = 5, 7$, Model 1 predicts higher survival probability than Model 2 and the two-stage approach. These differences could be due to different estimates of $\alpha_0, \dots, \alpha_6$ and γ_0, γ_1 in the survival sub-models of Models 1 and 2. At $r = 9$, Model 1 picked up the change in CD4 counts driven by the HAART initiation and the estimated CD4 trajectory started to increase, while the estimated trajectory from Model 2 still indicated a decreasing pattern. Thus, this leads to the higher predicted survival probabilities from model 1 compared with those from Model 2 at $r = 9$, which is easier to be seen at the probit scale in Figure S2. The number of longitudinal data points required to make reliable predictions will therefore depend on the true individual trajectory, because to make predictions, we linearly extrapolate from the last observed data point before the prediction time. The survival model with the time-varying covariate and the two-stage approaches give similar predictions as for Model 2 for $r = 3, 5$. For $r = 9$, the survival model with the time-varying covariate gives slightly higher predicted survival probabilities, possibly because of the large value of the last observed CD4 count by the cutoff time. Overall, our joint model can capture the nonlinear change in the individual longitudinal trajectory, which is helpful to provide more accurate predictions of conditional survival probabilities.

5. Simulation study

In this section, we perform a simulation study to evaluate the dynamic prediction performance of the proposed flexible joint model by comparisons with (i) a survival model using the longitudinal outcome as a time-varying covariate; (ii) a two-stage approach that uses empirical Bayes estimates of random effects, based on a linear mixed model with P-splines fitted to observed longitudinal data, in a subsequent
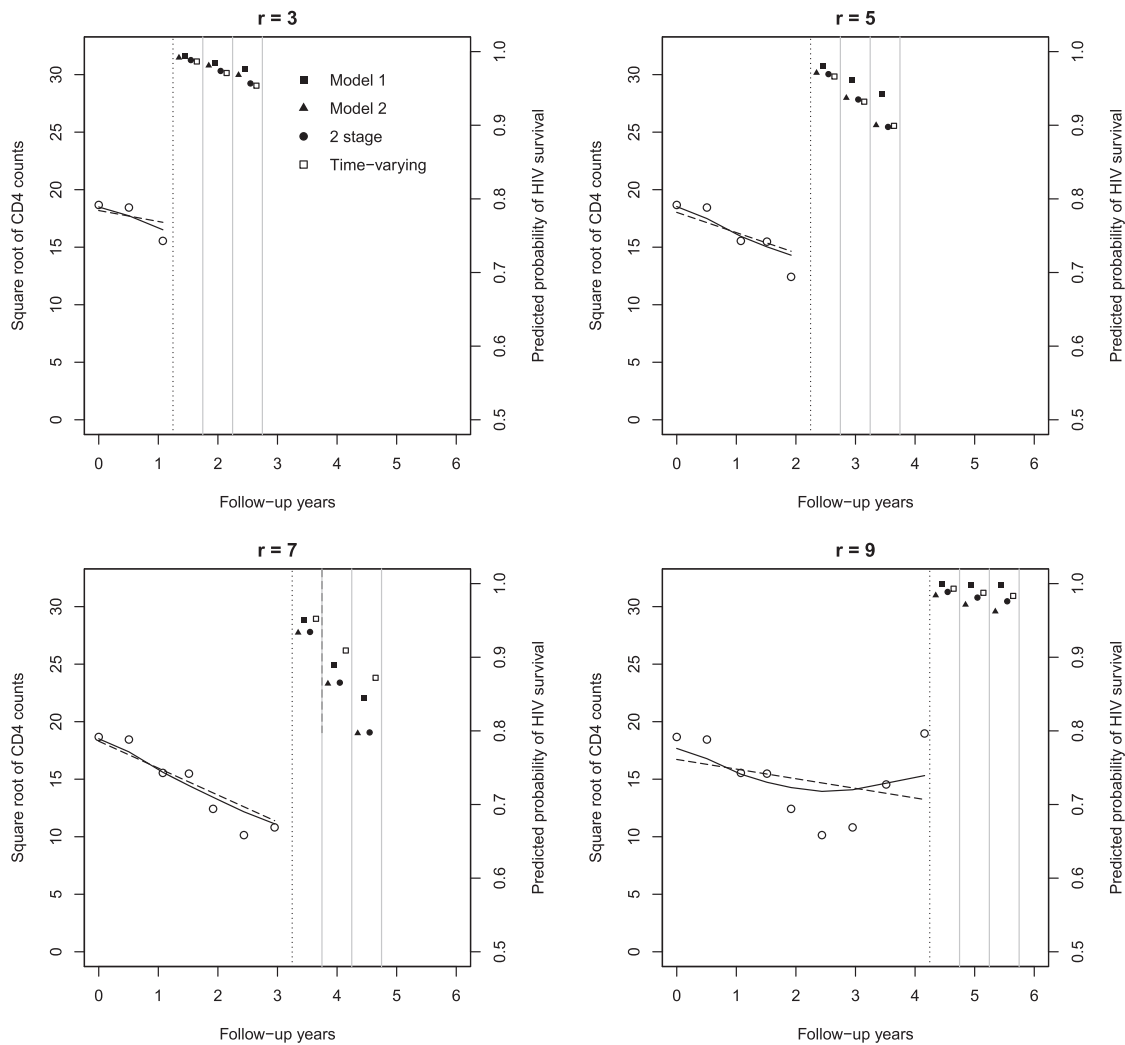
Figure 3. Left side of each panel: observed (standardized) square root CD4 counts and estimated individual longitudinal trajectory for patient 26 up to the prediction time $r = 3, 5, 7, 9$. The dotted lines are the cutoff time for the prediction. The solid and dashed dark lines are estimated individual longitudinal trajectories (using medians of 200 samples from the posterior of $\mathbf{b}_i$) based on Models 1 and 2, respectively. Right side of each panel: predicted conditional probabilities of HIV survival after the next one, two, and three time intervals. Squares represent predictions from Model 1 and triangles represent predictions from Model 2 (using medians of 200 samples from the posterior of $\mathbf{b}_i$).

survival model; (iii) a joint model with random intercept and slope only; and (iv) a joint model with cubic splines and three internal knots.

Except for the survival model with CD4 counts as time-varying covariate, the same survival sub-model is specified for the three joint models as well as in the two-stage approach. We will use the 'gold standard' estimator of $\pi_i(s \mid r)$ with the true (i.e., simulated) values for random effects and true values for the parameters and evaluate the dynamic prediction performance as a function of the prediction time r and also the prediction window $\Delta t = s - r$. The setup for the simulation is motivated by the HERS data, and details are given in Section S3.

The main conclusions drawn from the simulation study are as follows. First, the proposed joint model with P-splines outperforms the other two joint models overall (aggregated over all prediction times). Second, the two-stage approach performs only slightly worse than the joint model based on P-splines in dynamic predictions, although their performances in parameter estimation are very different. The survival model using the longitudinal outcome as a time-varying covariate performs the worst among all approaches. Third, depending on the shape of the true longitudinal trajectories and the prediction time r, the joint model with random intercept and slope and the joint model with cubic splines can perform sim-

ilarly to the proposed joint model. Finally, the joint model with cubic splines performs the worst among the three joint models, especially at later prediction times r. More detailed results can be found in the Supporting Information.

Overall, our simulation results show that our flexible joint model had better dynamic prediction performance than all other approaches in comparison.

6. Discussion

In this paper, we developed a flexible joint model for longitudinal and time-to-event data with time-dependent random effects, aiming to improve dynamic predictions by allowing for more flexible modeling of the longitudinal process. Our simulation results demonstrate that our flexible joint model with P-splines and truncated linear basis outperformed other existing approaches in terms of parameter estimation and dynamic predictions. Moreover, in practice, it is straightforward to implement dynamic predictions based on our model because the posterior distribution of the random P-spline coefficients is a multivariate skew-normal distribution.

The poor performance of the joint model with cubic splines in the simulations might be explained by the fact that splines with polynomial bases tend to perform erratically beyond the boundary knot and extrapolation can be dangerous [16, Chapter 5]. For this reason, natural cubic splines are often used to add a linear constraint beyond the boundary knot. When making dynamic predictions, however, at the time of prediction, we must extrapolate an individual's trajectory beyond the current observed data for that individual. Thus, using natural cubic splines in a joint model is not very helpful if the linear constraint is only applied to the last boundary knot in the whole follow-up. Adding further flexibility to the cubic spline model in our simulation study would be very likely to exacerbate the extrapolation problem. Our joint model with P-splines and truncated linear basis not only offers model flexibility (compared with a joint model with random intercept and slope) but also alleviates the extrapolation problem beyond the prediction time (compared with a joint model with cubic splines). Note that this discussion is only applied to modeling individual longitudinal trajectories. For population-level longitudinal trajectories, other spline bases with better numerical properties could be used in our model. Overall, based on the findings from our simulation study, for dynamic prediction purpose only, we do not recommend using splines with polynomial bases for modeling individual longitudinal trajectories.

Interestingly, the two-stage approach using P-splines and truncated linear basis also performed very well in terms of dynamic predictions, although it produced biased parameter estimates in both longitudinal and survival sub-models. The bias–variance trade-off is well known for prediction problems [16, Chapter 7]. The two-stage approach and the joint model with cubic splines had similar biases in terms of parameter estimation. However, the complexity in the joint model with cubic splines also introduced more variance into the estimation, which was demonstrated by the larger variability in prediction errors for the joint model with cubic splines at later prediction times. Therefore, overall the joint model with cubic splines performed worse than the two-stage approach for dynamic predictions in our simulations. In practice, if dynamic predictions are of main interest, two-stage approaches can be applied without the complexity of fitting joint models. Note that the extrapolation problem for polynomial bases discussed earlier applied to the two-stage approach as well.

Barrett *et al.* investigated the impact of discretization of the time scale on the inferences of the longitudinal and survival sub-models [14]. Their simulation studies and analysis of special cases suggested that the parameter estimates were not greatly influenced by the discretization, in particular, the covariate effects in the longitudinal and survival sub-models. Moreover, Barrett *et al.* theoretically proved that there is no loss of information when the survival functions are linear between discrete time points [14]. In practice, often there exists a natural discrete time scale, for example, dropout at pre-specified measurement time points. For continuous-time dropout or other continuous time to event, a discretization that ensures approximate linearity is recommended.

Using a probit model for the discretized event time, our model benefits from the straightforward implementation of dynamic prediction of survival probabilities. The probit link used in the survival sub-model not only facilitates estimation but also naturally reflects the assumption that the discrete hazard of event occurrence depends on the *normally* distributed random effects that characterize underlying individual longitudinal trajectories. In other words, because we assume that the linear predictor in the survival sub-model is normally distributed, it seems natural to use the probit link to transform back to the discrete hazard (probability) scale. To interpret the covariate effects in the survival sub-model, we can present the results at the marginal survival probability scale to the subject matter experts.

Barrett *et al.* discussed the computational issues related to their joint model with time-dependent random effect [14], which are similar in our case as the computing time is also driven primarily by calculating the multivariate normal probabilities. The R package `mnormt` we used applies a non-Monte Carlo method to calculate multivariate normal probabilities up to 20 dimensions. Another R package `mvtnorm` uses faster quasi-Monte Carlo methods and can accommodate dimensions up to 1000 but with less accuracy. The development in this field will certainly help improve our estimation procedure.

A limitation of our proposed model is that due to discretization of the time scale, dynamic predictions can only be made in discrete time intervals as well. But given that prediction on patient prognosis is often made in discrete time in practice, for example, 6-year survival given that the patient is still alive at 5 years, this limitation should not be a major concern.

In our simulation study, we compared the dynamic prediction performance of three joint models, assuming that the main structure of the survival sub-model is correctly specified. However, in practice, it is important to realize that model specification or different parameterizations in the survival sub-model can lead to different prediction estimates for conditional survival probabilities. Rizopoulos [7, Chapter 7] compared dynamic prediction results from six joint models with different parameterizations in the survival sub-model and found that the predicted conditional survival probabilities showed considerable variability between the six parameterizations. In practice, the choice for parameterizations in the survival sub-model should be mainly driven by substantive knowledge. When it is not available, standard likelihood information criteria can be used to decide upon which joint model we should base the predictions [7].

Acknowledgements

References

1. Gould AL, Boye ME, Crowther MJ, Ibrahim JG, Quartey G, Micallef S, Bois FY. Joint modeling of survival and longitudinal non-survival data: current methods and issues. Report of the DIA Bayesian joint modeling working group. *Statistics in Medicine* 2015; **34**:2181–95.
2. Yu M, Taylor JMG, Sandler H. Individulized prediction in prostate cancer studies using a joint longitudinal–surival–cure model. *Journal of the American Statistical Association* 2008; **103**:178–187.
3. Proust-Lima C, Taylor JMG. Development and validation of a dynamic prognostic tool for prostate cancer recurrence using repeated measures of posttreatment PSA: a joint modeling approach. *Biostatistics* 2009; **10**:535–549.
4. Rizopoulos D. Dynamic predictions and prospective accuracy in joint models for longitudinal and time-to-event data. *Biometrics* 2011; **67**:819–829.
5. Taylor JMG, Park Y, Ankerst DP, Proust-Lima C, Williams S, Kestin L, Bae K, Pickles T, Sandler H. Real-time individual predictions of prostate cancer recurrence using joint models. *Biometrics* 2013; **69**:206–213.
6. Durbán M, Harezlak J, Wand MP, Carroll RJ. Simple fitting of subject-specific curves for longitudinal data. *Statistics in Medicine* 2005; **24**:1153–1167.
7. Rizopoulos D. *Joint Models for Longitudinal and Time-to-event Data, with Applications in R*. Chapman and Hall/CRC: Boca Raton, 2012.
8. Brown ER, Ibrahim JG, DeGruttola V. A flexible B-spline model for multiple longitudinal biomarkers and survival. *Biometrics* 2005; **61**:64–73.
9. Ding J, Wang J. Modeling longitudinal data with nonparametric multiplicative random effects jointly with survival data. *Biometrics* 2008; **64**:546–556.
10. Rizopoulos D, Ghosh P. A Bayesian semiparametric multivariate joint model for multiple longitudinal outcomes and a time-to-event. *Statistics in Medicine* 2011; **30**:1366–1380.
11. Crowther MJ, Abrams KR, Lambert PC. Flexible parametric joint modelling of longitudinal and survival data. *Statistics in Medicine* 2012; **31**:4456–4471.
12. Crowther MJ, Abrams KR, Lambert PC. Joint modeling of longitudinal and survival data. *The Stata Journal* 2013; **13**: 165–184.
13. Rizopoulos D, Verbeke G, Lesaffre E. Fully exponential Laplace approximations for the joint modelling of survival and longitudinal data. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 2009; **71**:637–654.
14. Barrett J, Diggle P, Henderson R, Taylor-Robinson D. Joint modelling of repeated measurements and time-to-event outcomes: flexible model specification and exact likelihood inference. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 2015; **77**:131–148.
15. Ruppert D, Wand MP, Carroll RJ. *Semiparametric Regression*. Cambridge University Press: Cambridge, 2003.

16. Hastie T, Tibshirani R, Friedman J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer: (New York), 2009.
17. Smith DK, Gardner LI, Phelps R, Hamburger ME, Carpenter C, Klein RS, Rompalo A, Schuman P, Holmberg SD. Mortality rates and causes of death in a cohort of HIV-infected and uninfected women, 1993–1999. *Journal of Urban Health: Bulletin of the New York Academy of Medicine* 2003; **80**:676–688.
18. Arnold BC. Flexible univariate and multivariate models based on hidden truncation. *Journal of Statistical Planning and Inference* 2009; **139**:3741–3749.
19. Eilers PHC, Marx BD. Flexible smoothing with B-splines and penalties (with discussion). *Statistical Science* 1996; **11**: 89–121.
20. Daniels MJ, Hogan JW. *Missing Data in Longitudinal Studies: Strategies for Bayesian Modeling and Sensitivity Analysis*, Monographs on Statistics and Applied Probability, vol. 101. Chapman & Hall/CRC: New York, 2008.

Supporting Information

Additional supporting information may be found online in the supporting information tab for this article.