Computational Statistics and Data Analysis 129 (2019) 14–29

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

journal homepage: www.elsevier.com/locate/csda

# Bayesian functional joint models for multivariate longitudinal and time-to-event data

Kan Li [a], Sheng Luo [b,*]

[a] Merck Research Lab, Merck & Co, 351 North Sumneytown Pike, North Wales, PA 19454, USA
[b] Department of Biostatistics and Bioinformatics, Duke University Medical Center, 2400 Pratt St, 7040 North Pavilion, Durham, NC 27705, USA

ARTICLE INFO

*Article history:*
Received 30 January 2018
Received in revised form 30 June 2018
Accepted 26 July 2018
Available online 16 August 2018

*Keywords:*
Longitudinal functional data
Joint modeling
Dynamic prediction
Alzheimer's disease

ABSTRACT

A multivariate functional joint model framework is proposed which enables the repeatedly measured functional outcomes, scalar outcomes, and survival process to be modeled simultaneously while accounting for association among the multiple (functional and scalar) longitudinal and survival processes. This data structure is increasingly common across medical studies of neurodegenerative diseases and is exemplified by the motivating Alzheimer's Disease Neuroimaging Initiative (ADNI) study, in which serial brain imaging, clinical and neuropsychological assessments are collected to measure the progression of Alzheimer's disease (AD). The proposed functional joint model consists of a longitudinal function-on-scalar submodel, a regular longitudinal submodel, and a survival submodel which allows time-dependent functional and scalar covariates. A Bayesian approach is adopted for parameter estimation and a dynamic prediction framework is introduced for predicting the subjects' future health outcomes and risk of AD conversion. The proposed model is evaluated by a simulation study and is applied to the motivating ADNI study.


## 1. Introduction

The growing public health threat posed by Alzheimer's disease (AD) has raised the urgency to discover and assess markers for the early detection of the disease. In this regard, a great deal of effort has been dedicated to building models for predicting AD based on a single marker, or a combination of multiple markers, which captures the heterogeneity among subjects and detects the disease progression of subjects at risk (Weiner et al., 2013). Since mild cognitive impairment (MCI) is often considered as a transitional stage to AD, MCI patients are usually enrolled as the target population for early prognosis and evaluating interventions (Petersen et al., 1999). Existing research has identified a number of biomarkers in predicting an individual's likelihood of converting to AD, as well as differences in biomarker values among MCI and AD individuals (Perrin et al., 2009; Schmand et al., 2010). It is widely acknowledged that magnetic resonance imaging (MRI) based measures of atrophy in key brain regions, such as the hippocampus, are predictive of progression from MCI to AD (Du, 2001; Frisoni et al., 2010). Although most of the current studies measure regional atrophy using a single volume-based value, some researchers (Apostolova et al., 2010; Qiu et al., 2009) demonstrated that the surface-based morphology analysis offers more advantages because this method studies patterns of subregion atrophy and produces detailed pointwise correlation between atrophy and cognitive decline. Li and Luo (2017b) proposed a functional joint model (FJM) that incorporates surface-based hippocampus measure as a functional predictor in the joint model of longitudinal and survival framework. They developed a dynamic prediction method and demonstrated that using such a functional predictor, in addition to other scalar markers, improves

* Corresponding author.
*E-mail address:* sheng.luo@duke.edu (S. Luo).

https://doi.org/10.1016/j.csda.2018.07.015
0167-9473/

predictive performance of the progression of MCI to AD (Li and Luo, 2017a). However, the proposed FJM only accommodates baseline imaging marker as a time-invariant function predictor. Since the imaging markers (e.g., hippocampus) from MRI, along with other neurocognitive markers, are often collected repeatedly in the studies of AD, it is of scientific interest to investigate the combined predictive performance of these repeatedly measured functional and scalar outcomes.

Several methods for the analysis of repeatedly measured functional outcome exist in the literature. One category of the methods is based on functional principal component analysis (FPCA), as well as its extension for multilevel FPCA by Di et al. (2009), longitudinal FPCA by Greven et al. (2010) and by Park and Staicu (2015). These methods modeled subject-specific deviations from a population mean by using low dimensional basis functions estimated from the empirical covariance matrix. However, they were inflexible to estimate the effect of covariates (e.g., age) on the functional outcome. Brumback and Rice (1998) and Guo (2002) proposed a function-on-scalar mixed effect model in which population level effects and individual level deviations were modeled by using penalized splines. Wavelet-based Bayesian functional mixed models were presented in Morris and Carroll (2006), which used a discrete wavelet transform of the observed functional data and modeled coefficients in the wavelet domain. Goldsmith and Kitago (2016) developed a Bayesian framework for penalized spline function-on-scalar regression, allowing the joint modeling of population level fixed effects, individual level random effects, and residual functions. However, these works focused on the statistical inference on longitudinal functional data without considering the survival process and not for prediction purpose.

Joint model is an appropriate framework to modeling longitudinal data and time-to-event data since it has potential to reduce parameter estimate bias, account for dropout in longitudinal studies, and enable the inclusion of longitudinal covariates (both scalar and functional) measured with error in time-to-event models (Tsiatis and Davidian, 2004; Henderson et al., 2000). Multivariate joint models have been well studied by considering multivariate continuous, binary, ordinal, or a mixture of different outcome types. Hickey et al. (2016) gave an excellent review of multivariate joint modeling research. However, no previous study investigates how to incorporate the longitudinal functional (high-dimensional) outcome in a multivariate joint model framework. To this end, we propose a novel joint model that incorporates the growing volume of repeatedly measured functional outcomes in the longitudinal–survival setting. Specifically, we develop a multivariate functional joint model (MFJM) that could simultaneously analyze a longitudinal functional outcome, a longitudinal scalar outcome, and a survival outcome. The principle of the MFJM is to define three type of submodels: (1) a functional mixed effect submodel for the longitudinal functional outcome, (2) a regular mixed effects submodel, or multiple regular mixed effects submodels, to describe the evolution of the longitudinal scalar outcome(s), and (3) a Cox submodel for the survival outcome which is linked with (1) and (2) using a common latent structure. The MFJM is flexible to account for the correlation between repeated measures and correlation among multiple outcomes. We estimate the coefficient functions in the functional regression using penalized spline approach and parameters are jointly estimated in a Bayesian framework.

Compared with the existing literature, we make two major contributions to both multivariate joint modeling and functional data analysis: (1) We propose a multivariate joint model considering both longitudinal functional and scalar outcomes. To the best of our knowledge, this paper is the first to model the repeatedly measured functional outcomes, scalar outcomes, survival process simultaneously while accounting for the associations among the processes. (2) We propose a dynamic prediction framework that provides accurate personalized predictions of disease risk and progression. We investigate the potential capability of the longitudinal functional outcome in improving the prediction of AD progression. Previous studies involving functional data mainly focused on model inference rather than prediction of risk and longitudinal outcome trajectories. These important predictive tools can provide valuable information to monitor each patient's disease progression and to make early decisions about targeted prevention and treatment selection.

The remainder of the article is organized as follows. In Section 2, we describe the motivating Alzheimer's Disease Neuroimaging Initiative (ADNI) study and the data structure. In Section 3, we discuss the multivariate functional joint model, Bayesian inference procedure, and dynamic prediction framework. In Section 4, we apply the proposed method to the motivating ADNI study. In Section 5, we conduct a simulation study to assess the performance of the method. Concluding remarks and discussion are presented in Section 6.

## 2. A motivating clinical study

The methodology development is motivated by Alzheimer's Disease Neuroimaging Initiative (ADNI) study. The primary goal of the study is to test whether serial magnetic resonance imaging (MRI), positron emission tomography (PET), cerebrospinal fluid (CSF) markers, and neuropsychological assessments can be combined to measure the progression of AD. The phase one of the ADNI study (ADNI-1) recruited more than 800 adults, of which about 200 cognitively normal individuals, 400 mild cognitive impairment (MCI) patients, and 200 early AD patients. Participants were reassessed at 6, 12, 18, 24 and 36 months, and additional follow-ups were conducted annually as part of ADNI-2. At each visit, various neuropsychological assessments, brain image, and clinical measures were collected. Detailed information about the ADNI study procedures, including participant inclusion and exclusion criteria and complete study protocol can be found at http://www.adni-info.org.

MCI is commonly considered as a transitional stage between normal cognition and Alzheimer's disease and used as the target population for evaluating prognosis and early treatment. To this end, our analysis focuses on 355 MCI patients in the ADNI-1 study without missing data in covariates of interests, and we consider time from baseline to AD diagnosis among MCI patients to be the survival event of interest. In the ADNI-1 study, the 355 MCI patients were followed up for a mean

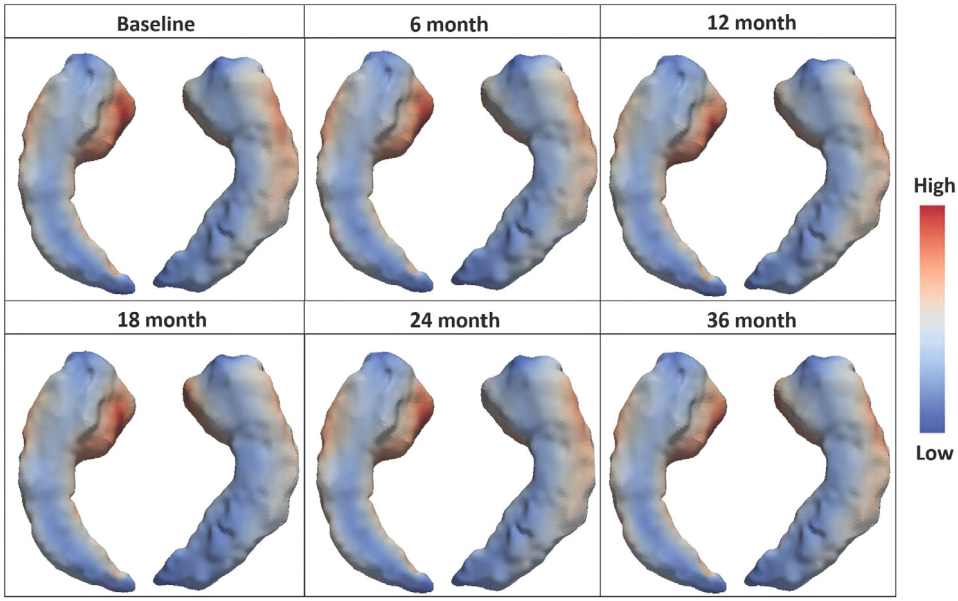


**Fig. 1.** The longitudinal profile of surface-based hippocampal images of one MCI patient: hippocampal radial distances are denoted by colors.

of 3.2 years (SD 2.6; range 0.4–9.3) before AD diagnosis or censoring. Among them, 180 patients were diagnosed with AD (survival event) and 175 had stable MCI over a mean follow-up period of 2.3 years and 4.2 years, respectively.

Longitudinal AD Assessment Scale-Cognitive (ADAS-Cog) score and Hippocampal volume (HV) were reported to be the strongest predictor of AD progression in the cognitive and imaging domains, respectively (Li et al., 2017). However, when the high-dimensional MRI data are aggregated to single volume data such as HV, enormous information is lost (Qiu et al., 2009), as the more recent surface-based morphology analysis (based on the longitudinal changes of cortical thickness in thousand of vertices) provides crucial disease progression information for early detection of AD (Huang et al., 2017). In this study, we adopt the surface-based analysis of imaging data which retains more information about Hippocampus morphology. In the surface-based analysis, the hippocampus is modeled as a surface model which is a mesh of triangles. Each triangle is known as a face and the place where the corners of the triangles meet is called a vertex. The coordinate of each vertex is determined during image processing and allows one to compute many morphometric measures, e.g., hippocampal radial distance (HRD). Fig. 1 illustrates the longitudinal profile of the surface-based hippocampus images (mapped on a three-dimensional hippocampus template) of one MCI patient at different visits. The colors represent the hippocampal radial distance (HRD) which measures the distance from the medial core to each point on the surface (referred to as vertex) and reflects the hippocampal cortical thickness. As AD progresses and the hippocampus atrophies, the radial distance of some subfields shrinks. It has been shown that the baseline vertex-based HRD is predictive of time of MCI-to-AD as a functional predictor (Li and Luo, 2017a). In this paper, we propose a Bayesian personalized prediction model based on a multivariate functional joint model (MFJM) of longitudinal ADAS-Cog 11 score as a scalar predictor, longitudinal vertex-based HRD as a functional predictor, and the time to AD diagnosis.

The image processing procedure is detailed in the Web Supplement. We first extract the hippocampal surfaces (left and right) from original MRI scans (Step 1 in Web Figure 1) using FIRST (Patenaude et al., 2011), an integrated surface analysis tool developed as part of the FSL library (Jenkinson et al., 2012). The surfaces are then conformally mapped to a two-dimensional (2D) rectangle plane, in the form of matrix, to form two feature images (Step 2). We then register each feature image (patient and visit) to a common template and calculate the hippocampal radial distance (HRD) of each vertex to the predefined medical core, which represents the hippocampal cortical thickness (Step 3). These steps account for the spatial information and image smoothing. Then the HRD values on the 2D image matrices are aggregated over the $y$-axis of the image into a one-dimensional (1D) image vector such that the corresponding HRDs of the vertices are represented as a 1D functional data (denoted by $y_i(s, t_{ij})$ for subject $i$ visit $j$) defined on domain $\mathcal{S}$ (Step 4). Each point in the image vector domain (i.e., $\mathcal{S}$) corresponds to a coordinate on $x$-axis of the 2D hippocampal image matrix. It was revealed that left hippocampus atrophy was associated with delayed verbal memory (Chen et al., 2010), where the delayed verbal memory was one of the important predictors for determine whether a subject was a MCI converter or not (Gomar et al., 2011). Thus, our analysis focus on the surface morphology and HRD of the left hippocampus.

## 3. Methods

### 3.1. Multivariate functional joint model (MFJM) framework

In the context of clinical trials and observational studies, for each subject $i$ ($i = 1, \ldots, I$) at visit $j$ ($j = 1, \ldots, J_i$) and on a 1D domain $s \in [0, S_{max}] = \boldsymbol{S}$, we observe data $\{y_{ij}^*, y_{ij}(s), \boldsymbol{x}_{ij}\}$, where $y_{ij}^* = y_i^*(t_{ij})$ is a scalar response observed at time $t_{ij}$ from the study onset, $y_{ij}(s) = y_i(s, t_{ij})$ is a functional response curve observed at time $t_{ij}$ on domain $\boldsymbol{S}$, and $\boldsymbol{x}_{ij} = [x_{ij1}, \ldots, x_{ijP}]$ is a $P$-dimensional scalar covariates vector. The domain of the functional response $\boldsymbol{S}$ is not the same as the time domain $t$, over which the survival event is followed. We propose a longitudinal submodel to describe the evolution of the scalar outcome over time. The model is represented as

$$y_i^*(t_{ij}) = m_i^*(t_{ij}) + \epsilon_{ij}^*, \quad m_i^*(t_{ij}) = \beta_0 + t_{ij}\beta_t + \sum_p^P x_{ijp}\beta_p + \boldsymbol{V}_R(t_{ij})\boldsymbol{\zeta} + b_i^*, \tag{1}$$

where $m_i^*(t_{ij})$ is the unobserved true value of the scalar longitudinal outcome at time $t_{ij}$, $\beta_0$ is the intercept, $\beta_t$ is the change of scalar outcome overtime, $\beta_p$'s are the regression coefficients. To allow additional flexibility and smoothness in modeling the effects of some covariates, we adopt a smooth time function $\boldsymbol{V}_R(t)\boldsymbol{\zeta} = \sum_{r=1}^R \zeta_r(t-\kappa_r)_+$ using the truncated power series spline basis expansion $\boldsymbol{V}_R(t) = \{(t-\kappa_1)_+, \ldots, (t-\kappa_R)_+\}$, where $\boldsymbol{\zeta} = [\zeta_1, \ldots, \zeta_R]^\top$ are the spline coefficients, $\boldsymbol{\kappa} = \{\kappa_1, \ldots, \kappa_R\}$ are the knots, and $(t-\kappa_r)_+ = t - \kappa_r$ if $t > \kappa_r$ and 0 otherwise. We consider a sufficient large number of knots that can ensure the desired flexibility and we select the knot location to have sufficient subjects between adjacent knots. The choice of knots is important to obtain a well fitted model and penalizing the spline coefficients to constrain their influence could help to avoid overfitting (Ruppert et al., 2003). The random intercepts $b_i^*$ are independent and identically distributed (iid), and the measurement errors $\epsilon_{ij}^* \sim N(0, \sigma_{\epsilon^*}^2)$ are independent from $b_i^*$. The inclusion of covariate-specific random effects as a random slope is a direct extension of model (1).

We assume that the functional response $y_i(s, t_{ij})$ is linear in time, which is a direct extension of the linear assumption in the scalar model. The longitudinal functional submodel is defined as

$$y_i(s, t_{ij}) = m_i(s, t_{ij}) + \epsilon_{ij}(s), \quad m_i(s, t_{ij}) = B_0(s) + t_{ij}B_t(s) + \sum_p^P x_{ijp}B_p(s) + b_i(s), \tag{2}$$

in which $m_i(s, t_{ij})$ is the unobserved true value of the longitudinal functional outcome at time $t_{ij}$ over domain $\boldsymbol{S}$, $B_0(s)$ is the overall mean function, $B_t(s)$ and $B_p(s)$'s are fixed effect coefficient functions corresponding to time $t$ and the scalar covariates $\boldsymbol{x}_{ij}$. The random intercept function $b_i(s)$ for subject $i$ represents the subject-specific effect, and $\epsilon_{ij}(s)$ is a white noise error process with covariance $cov\{\epsilon_{ij}(s), \epsilon_{ij}(s')\} = \sigma_\epsilon^2$ if $s = s'$ and 0 otherwise. We assume that the random function $b_i(s)$ are iid, the error process $\epsilon_{ij}(s)$ are iid and are independent from $b_i(s)$. For identifiability we require that $b_i(s)$ comprises solely the random deviation that is specific to the subject; any repeated visit-specific deviation is viewed as part of $\epsilon_{ij}(s)$. As in traditional mixed models, the inclusion of "random slope functions" would allow subject-specific impacts of changing covariate levels and should be considered in future applications.

The event history is recorded for each subject $i$ with observed event time $T_i = min(T_i^*, C_i)$ and the event indicator $\delta_i = I(T_i^* \leq C_i)$, where $T_i^*$ and $C_i$ are the true event time and censoring time, respectively. The survival submodel is

$$h_i(t) = h_0(t)\exp\{\boldsymbol{w}_i^\top\boldsymbol{\gamma} + \alpha^* m_i^*(t) + \int_{\boldsymbol{S}} \alpha(s)m_i(s, t)ds\}, \tag{3}$$

where $h_0(t)$ is the baseline hazard function, and $\boldsymbol{w}_i$ is a vector of time-independent covariates with regression coefficient vector $\boldsymbol{\gamma}$. The association parameter $\alpha^*$ quantifies the strength of correlation between the unobserved true longitudinal function $m_i^*(t)$ and the event hazard at the same time point $t$, and the association function $\alpha(s)$ quantifies the correlation between the unobserved true longitudinal function $m_i(s, t)$ and the event hazard at the time point. In this paper, we assume a constant functional parameter $\alpha(s) \equiv \alpha$ for identifiability and discuss the case that $\alpha(s)$ varies over the domain of $\boldsymbol{S}$. We implicitly assume that the risk for an event at time $t$ depends on the unobserved true value of the longitudinal outcomes at the same time point in Model (3). However, other functional forms for the association structure such as time-dependent slopes or cumulative effects of $m_i^*(t)$ and $m_i(s, t)$ can also be included in the survival submodel. Models (1), (2), and (3) consist of the multivariate functional joint model (MFJM) framework.

To modeling the longitudinal functional data, we adopt the functional mixed effect model and expand the random intercept function $b_i(s)$ using a functional principal component analysis (FPCA) approach. FPCA is a dimensionality reduction tool for functional data which leads to low dimensional projection basis (eigenfunction) and makes analyzing data easier. Specifically, we first express the random intercept function $b_i(s)$ in model (2) using the Karhunen–Loève decomposition. The spectral decomposition of the covariance function of $b_i(s)$'s is given by $\Sigma^{(b)}(s, s') = \sum_{k=1}^\infty \lambda_k\phi_k(s)\phi_k(s')$, where $\lambda_1 \geq \lambda_2 \geq \cdots \geq 0$ are non-increasing eigenvalues and $\phi_k(s)$'s are the corresponding eigenfunctions. The Karhunen–Loève expansion of $b_i(s)$ is $b_i(s) = \sum_{k=1}^\infty \xi_{ik}\phi_k(s)$, where the functional principal component (FPC) scores $\xi_{ik}$ are uncorrelated random variables

with mean zero and variance $\lambda_k$. In practice, we adopt a truncated approximation for $b_i(s)$ given by $b_i(s) \approx \sum_{k=1}^{K_\phi} \xi_{ik}\phi_k(s)$. Thus $y_i(s, t_{ij})$ is written as

$$y_i(s, t_{ij}) \approx B_0(s) + t_{ij}B_t(s) + \sum_{p}^{P} x_{ijp}B_p(s) + \sum_{k=1}^{K_\phi} \xi_{ik}\phi_k(s) + \epsilon_{ij}(s).$$

The number of eigenfunctions for random intercept function $K_\phi$ is pre-specified fixed constants. A sufficiently large value should be chosen for $K_\phi$ to capture the variation in the random functions, and sensitivity to the choices should be assessed. However, selecting the number of eigenfunctions larger than necessary leads to increased computing burden (Goldsmith and Kitago, 2016). For computing efficiency, we assume that the correlation between $y_i^*(t_{ij})$ and $y_i(s, t_{ij})$ is manifested by the correlation between $b_i^*$ and the first elements in $\boldsymbol{\xi}_i = [\xi_{i1}, \ldots, \xi_{iK_\phi}]$, and $\boldsymbol{b}_i = [b_i^*, \boldsymbol{\xi}_i] \sim MVN(\mathbf{0}, \boldsymbol{\Sigma})$, where

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_{b^*}^2, & \rho\sigma_{b^*}\sqrt{\lambda_1}, & \ldots, & 0 \\ \rho\sigma_{b^*}\sqrt{\lambda_1}, & \lambda_1, & \ldots, & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0, & 0, & \ldots, & \lambda_{K_\phi} \end{bmatrix}.$$

We may also estimate the correlation between the scalar random effects and all FPC components via an covariance matrix whose off-diagonal elements are nonzero. Such a covariance matrix may provide a full representation of the correlation between the mixed outcomes. However, as the computational burden increases dramatically as the covariance matrix gets more complex, we have to consider the trade-off among modeling flexibility, estimation accuracy, and computation.

In practice, functional outcomes are not truly functions but are observed on a finite grid of length $M$ that cover the domain $\boldsymbol{S}$, i.e., $\{s_1, \ldots, s_M\}$. Let $\boldsymbol{B}(s)$ be the $M \times (P+2)$ matrix with columns containing $B_0(s)$, $B_t(s)$, and $B_p(s)$'s evaluated on the finite grid and let $\boldsymbol{\phi}(s)$ be a $M \times K_\phi$ matrix with columns containing eigenfunctions $\phi_k(s)$. We express the coefficient function and the eigenfunction in each column of $\boldsymbol{B}(s)$ and $\boldsymbol{\phi}(s)$ in terms of a known cubic $B$-spline with equally spaced knots, which leads to a $M \times K_\psi$ matrix $\boldsymbol{\psi}(s) = [\psi_1^\top(s), \ldots, \psi_{K_\psi}^\top(s)]$ with basis functions as columns. For example, the coefficient function $B_0(s) = \sum_{l=1}^{K_\psi} B_{0l}\psi_l(s) = [\boldsymbol{B}_0\boldsymbol{\psi}^\top(s)]^\top$ and the eigenfunction $\phi_k(s) = \sum_{l=1}^{K_\psi} B_{\phi_k l}\psi_l(s) = [\boldsymbol{B}_{\phi_k}\boldsymbol{\psi}^\top(s)]^\top$, where $\boldsymbol{B}_0$ and $\boldsymbol{B}_{\phi_k}$ are row vectors of spline coefficients $B_{0l}$ and $B_{\phi_k l}$, respectively. Other spline bases could be used, but the parameters of the $B$-spline have good mixing properties in the context of Bayesian posterior simulation (Crainiceanu and Goldsmith, 2010) and $B$-spline is widely used in functional data analysis literature for its flexibility (Morris, 2015). Let $\boldsymbol{B}_t$ and $\boldsymbol{B}_k$'s have the same meaning as $\boldsymbol{B}_0$, then $\boldsymbol{B} = [\boldsymbol{B}_0^\top, \boldsymbol{B}_t^\top, \boldsymbol{B}_1^\top, \ldots, \boldsymbol{B}_P^\top]^\top$ and $\boldsymbol{B}_\phi = [\boldsymbol{B}_{\phi_1}^\top, \ldots, \boldsymbol{B}_{\phi_{K_\phi}}^\top]^\top$ denote $(P+2) \times K_\psi$ and $K_\phi \times K_\psi$ matrices, respectively, whose rows are spline coefficients for $\boldsymbol{B}(s)$ and $\boldsymbol{\phi}(s)$. Therefore, the coefficient functions are written as $\boldsymbol{B}(s) = [\boldsymbol{B}\boldsymbol{\psi}^\top(s)]^\top$, the eigenfunctions are $\boldsymbol{\phi}(s) = [\boldsymbol{B}_\phi\boldsymbol{\psi}^\top(s)]^\top$, and the random intercept function is $b_i(s) = \boldsymbol{\xi}_i(\boldsymbol{\phi}(s))^\top = \boldsymbol{\xi}_i\boldsymbol{B}_\phi\boldsymbol{\psi}^\top(s)$, with $\boldsymbol{\xi}_i$ being the row vector of FPC scores for the random intercept function. For the choice of number of basis functions $K_\psi$ for $B$-spline, we refer to Ruppert (2002) and choose them large (e.g., 10) to capture the complexity in coefficient functions. We adopt penalization technique to prevent overfitting and induce smoothness in the resulting coefficient functions. Thus the functional longitudinal submodel is rewritten as

$$\begin{aligned} y_i(s, t_{ij}) &\approx m_i(s, t_{ij}) + \epsilon_{ij}(s), \text{ where} \\ m_i(s, t_{ij}) &\approx \boldsymbol{B}_0\boldsymbol{\psi}^\top(s) + t_{ij}\boldsymbol{B}_t\boldsymbol{\psi}^\top(s) + \sum_{p}^{P} x_{ijp}\boldsymbol{B}_p\boldsymbol{\psi}^\top(s) + \boldsymbol{\xi}_i\boldsymbol{B}_\phi\boldsymbol{\psi}^\top(s). \end{aligned} \tag{4}$$

We assume a constant associate function $\alpha(s) \equiv \alpha$, and thus the survival submodel is

$$h_i(t) = h_0(t)\exp\{\boldsymbol{w}_i^\top\boldsymbol{\gamma} + \alpha^* m_i^*(t) + \alpha\int_{\boldsymbol{S}} m_i(s, t)ds\}. \tag{5}$$

The integral in the survival submodel is computed by numeric integration.

Let $\boldsymbol{\theta} = \left[\boldsymbol{\beta}^\top, \boldsymbol{\zeta}^\top, \boldsymbol{B}^\top, (\boldsymbol{\xi}_i : i = 1, \ldots, I), (\lambda_k : k = 1, \ldots, K_\phi), \rho, \boldsymbol{B}_\phi^\top, \boldsymbol{\gamma}, \alpha^*, \alpha, \sigma_{b^*}^2, \sigma_{\epsilon^*}^2, \sigma_\epsilon^2, \boldsymbol{\theta}_{h_0^*}^\top\right]^\top$ be the unknown parameter vector to be estimated, where $\boldsymbol{\beta} = [\beta_0, \beta_t, \beta_1, \ldots, \beta_p]^\top$, and vector $\boldsymbol{\theta}_{h_0^*}$ denotes the parameters in the baseline hazard function $h_0^*(\cdot)$. The observed data $y_i^*(t_{ij})$, $y_i(s, t_{ij})$, $\boldsymbol{x}_{ij}$, $t_{ij}$, $\boldsymbol{w}_i$ for $i = 1, \ldots, I$ and $j = 1, \ldots, J_i$ are known; as well as the cubic $B$-spline basis functions $\boldsymbol{\psi}(s)$, which can be generated using the `bs` function in the `R` package `spline`.

The conditional likelihood from the longitudinal scalar data $\boldsymbol{y}_i^* = [y_{i1}^*, \ldots, y_{iJ_i}^*]^\top$ is

$$p(\boldsymbol{y}_i^*|\boldsymbol{\theta}, \boldsymbol{b}_i) = (2\pi\sigma_{\epsilon^*}^2)^{-J_i/2}\exp\left\{-\frac{1}{2\sigma_{\epsilon^*}^2}\sum_{j=1}^{J_i}\left[y_{ij} - \left(\beta_0 + t_{ij}\beta_t + \sum_{p}^{P} x_{ijp}\beta_p + \sum_{r=1}^{R}\zeta_r(t-\kappa_r)_+ + b_i^*\right)\right]^2\right\}.$$

The conditional likelihood for the functional longitudinal data $\boldsymbol{y}_i(s) = [y_{i1}^\top(s), \ldots, y_{ij_i}^\top(s)]^\top$ is

$$p(\boldsymbol{y}_i(s)|\boldsymbol{\theta}, \boldsymbol{b}_i) = |2\pi\sigma_\epsilon^2 \boldsymbol{I}_{s\times s}|^{-J_i/2} \exp\left\{-\frac{1}{2}\operatorname{tr}\left([\boldsymbol{y}_i(s) - \boldsymbol{m}_i(s)]^\top [\boldsymbol{y}_i(s) - \boldsymbol{m}_i(s)](\sigma_\epsilon^2 \boldsymbol{I}_{s\times s})^{-1}\right)\right\},$$

where $\boldsymbol{m}_i(s) = [m_i^\top(s, t_{i1}), \ldots, m_i^\top(s, t_{ij_i})]^\top$, $|\cdot|$ is the determinant of a matrix and tr is the trace of a matrix. The density function of the random effects $\boldsymbol{b}_i$ is $p(\boldsymbol{b}_i|\boldsymbol{\theta}) = (2\pi)^{-(K_{\boldsymbol{\phi}}+1)/2}|\boldsymbol{\Sigma}|^{-1/2} \exp(-\frac{1}{2}\boldsymbol{b}_i^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{b}_i)$, where $(K_{\boldsymbol{\phi}} + 1)$ is the dimension of the covariance matrix $\boldsymbol{\Sigma}$. The conditional likelihood from the survival data is

$$p(T_i, \delta_i, |\boldsymbol{\theta}, \boldsymbol{b}_i) = h_i(T_i|\boldsymbol{\theta}, \boldsymbol{b}_i)^{\delta_i} S_i(T_i|\boldsymbol{\theta}, \boldsymbol{b}_i) = h_i(T_i|\boldsymbol{\theta}, \boldsymbol{b}_i)^{\delta_i} \exp\Big[-\int_0^{T_i} h_i(t|\boldsymbol{\theta}, \boldsymbol{b}_i)dt\Big],$$

where $h_i(T_i|\boldsymbol{\theta}, \boldsymbol{b}_i) = h_0^*(T_i)\exp\left\{\boldsymbol{w}_i^\top \boldsymbol{\gamma} + \alpha^* m_i^*(T_i) + \alpha \int_{\boldsymbol{S}} m_i(s, T_i)ds\right\}$, and function $h_0^*(\cdot)$ can be approximated by a piecewise-constant function or a $B$-spline function.

Under the local independence assumption (i.e., conditional on the random effects vector $\boldsymbol{b}_i$, all components in $\boldsymbol{y}_i^*$, $\boldsymbol{y}_i(s)$, and $T_i$ are independent), the joint likelihood function is

$$L(\boldsymbol{\theta}) = \prod_{i=1}^{I} p(\boldsymbol{y}_i^*, \boldsymbol{y}_i(s), T_i, \delta_i|\boldsymbol{\theta}) = \prod_{i=1}^{I} \int p(\boldsymbol{y}_i^*|\boldsymbol{\theta}, \boldsymbol{b}_i)p(\boldsymbol{y}_i(s)|\boldsymbol{\theta}, \boldsymbol{b}_i)p(T_i, \delta_i, |\boldsymbol{\theta}, \boldsymbol{b}_i)p(\boldsymbol{b}_i|\boldsymbol{\theta})d\boldsymbol{b}_i. \tag{6}$$

### 3.2. Bayesian inference

For model estimation, we propose a Bayesian modeling approach based on Markov Chain Monte Carlo (MCMC) posterior simulations, which provides a flexible way for statistical inference. The Bayesian approach has a number of advantages and has been previously exploited in the univariate joint modeling framework (Faucett and Thomas, 1996) and functional regression (Crainiceanu et al., 2009). Liu and Li (2016) compared the performance of Bayesian approaches to classical frequentist (maximum likelihood) approaches under multivariate joint model framework, demonstrating superiority of the Bayesian methods with respect to bias, root-mean square error, and coverage.

We use vague priors on all elements in parameter vector $\boldsymbol{\theta}$. Specifically, the prior distributions of parameters $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$, $\alpha^*$, $\alpha$ are $N(0, 100)$, and Inverse-Gamma$(0.01, 0.01)$ for variance parameters $\sigma_{\epsilon^*}^2$ and $\sigma_\epsilon^2$. We impose smoothness on coefficient function estimates through the prior specification on spline coefficients $\boldsymbol{B}$ and $\boldsymbol{B}_\phi$, and assume the following priors for the columns of the matrices:

$$\begin{aligned}
\boldsymbol{B}_k &\sim N(0, \sigma_k^2 \boldsymbol{Q}^{-1}), \quad \text{for } 1 \le k \le (P+2),\\
\boldsymbol{B}_{\phi_k} &\sim N(0, \sigma_{\phi_k}^2 \boldsymbol{Q}^{-1}), \quad \text{for } 1 \le k \le K_{\boldsymbol{\phi}},
\end{aligned}$$

where $\boldsymbol{Q}$ is a pre-specified $K_{\boldsymbol{\psi}} \times K_{\boldsymbol{\psi}}$ penalty matrix enforces smoothness through the connection between Bayesian priors and quadratic penalization (Ruppert et al., 2003). As suggested in Goldsmith and Kitago (2016), we use $\boldsymbol{Q} = \mu\boldsymbol{Q}_0 + (1-\mu)\boldsymbol{Q}_2$ where $\boldsymbol{Q}_0$ and $\boldsymbol{Q}_2$ are zeroth- and second-order derivative penalty matrices, with the upper left parts as

$$\boldsymbol{Q}_0 = \boldsymbol{\psi}(s)\begin{bmatrix} 1 & 0 & 0 & 0 & \ldots \\ 0 & 1 & 0 & 0 & \ldots \\ 0 & 0 & 1 & 0 & \ldots \\ 0 & 0 & 0 & 1 & \ldots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}\boldsymbol{\psi}^\top(s) \quad \text{and} \quad \boldsymbol{Q}_2 = \boldsymbol{\psi}(s)\begin{bmatrix} 1 & -2 & 1 & 0 & 0 & 0 & \ldots \\ -2 & 5 & -4 & 1 & 0 & 0 & \ldots \\ 1 & -4 & 6 & -4 & 1 & 0 & \ldots \\ 0 & 1 & -4 & 6 & -4 & 1 & \ldots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}\boldsymbol{\psi}^\top(s),$$

where $\boldsymbol{\psi}(s)$ is the cubic $B$-spline evaluation matrix defined previously. Selecting $0 < \mu \le 1$ balance the universal shrinkage encoded in $\boldsymbol{Q}_0$ and the smoothness constraint of $\boldsymbol{Q}_2$, while ensuring $\boldsymbol{Q}$ is positive definite and priors are proper. In the simulation and real data analysis we set $\mu = 0.1$ and sensitivity analyses have indicated robustness to the choice of $\mu$ in the analysis. We use random walk prior of Lang and Brezger (2004) on the spline coefficients $\zeta_r$, for $r = 1, \ldots, R$, for smoothing and penalization. Specifically, we use a first order random prior distribution for $\zeta_{r+1} \sim N(\zeta_r, \sigma_\zeta^2)$, for $r = 1, \ldots, R-1$, where $\zeta_1$ is treated as a fixed unknown parameter. The variance component $\sigma_k^2$, $\sigma_{\phi_k}^2$, and $\sigma_\zeta^2$ are assigned Inverse-Gamma$(0.01, 0.01)$ as prior distribution. The parameters $\sigma_{b^*}^2$ and $\lambda_k$ in the covariance matrix $\Sigma$ are assigned Inverse-Gamma$(0.01, 0.01)$ prior distribution, and correlation coefficient $\rho$ is assigned Uniform$(-1, 1)$.

The model fitting is performed in `Stan` by specifying the full likelihood function and the prior distributions of all unknown parameters. `Stan` adopts a No-U-Turn sampler (NUTS), which is an extension to Hamiltonian Monte Carlo (HMC) that avoids random walk behavior by using the gradient of the log-posterior and eliminates the need to set a number of steps that required in HMC (Neal et al., 2011). NUTS uses a recursive algorithm to build a set of likely candidate points that spans a wide swath of the target distribution, stopping automatically when it starts to double back and retrace its steps (Hoffman and Gelman, 2014). Empirically, NUTS offers faster convergence and parameter space exploration compared with other MCMC algorithms such as Gibbs sampler. We use the history plots and view the absence of apparent trend in the plot as evidence of convergence. In addition, we use the Gelman–Rubin diagnostic to ensure the scale reduction $\hat{R}$ of all parameters

are smaller than 1.1 (Gelman et al., 2013). After fitting the model to the training dataset (the dataset used to build the model) using Bayesian approaches, we obtain $D$ (e.g., $D = 5000$ after burn-in) samples for the parameter vector denoted by $\{\boldsymbol{\theta}^{(d)}, d = 1, \ldots, D\}$. All estimations can then be obtained by calculating simple summaries (e.g., mean, variance, quantiles) of the posterior distributions of $D$ samples $\{\boldsymbol{\theta}^{(d)}, d = 1, \ldots, D\}$. Based on the estimated coefficient vector $\widehat{\boldsymbol{B}}$ (posterior mean), the estimated coefficient function is calculated by $\widehat{\boldsymbol{B}}(s) = [\widehat{\boldsymbol{B}}\boldsymbol{\psi}^\top(s)]^\top$. The Bayesian approach allows for the easy construction of posterior credible intervals for the coefficient function $B(s)$ as $[\widehat{q}_{B,0.025}(s), \widehat{q}_{B,0.975}(s)]$, where $\widehat{q}_{B,u}(s)$ is the $u$-quartile of the MCMC samples $\boldsymbol{B}(s)^{(d)} = [\boldsymbol{B}^{(d)}\boldsymbol{\psi}^\top(s)]^\top$, $d = 1, \ldots, D$. To facilitate easy reading and implementation of the proposed multivariate functional joint model, we provide a sample `Stan` code in the Web Supplement.

### 3.3. Dynamic prediction framework

We next illustrate the dynamic prediction framework based on the proposed model. Given a new subject $N$'s outcome histories $\boldsymbol{y}_N^{\{t\}} = \{y_N^*(t_{Nj}), y_N(s, t_{Nj}); 0 \le t_{Nj} \le t\}$ and covariates $\boldsymbol{X}_N^{\{t\}} = \{\boldsymbol{x}_N(t_{Nj}), \boldsymbol{w}_N; 0 \le t_{Nj} \le t\}$ up to time $t$, and $\delta_N = 0$ (no event), we want to predict the personalized scalar outcome $y_N^*(t')$ and functional outcome $y_N(s, t')$ at a future time point $t' > t$ (e.g., $t' = t + \Delta t$), as well as the conditional probability of event-free or survival at time $t'$, denoted by $\pi_N(t'|t) = P(T_N^* \ge t'|T_N^* > t, \boldsymbol{y}_N^{\{t\}}, \boldsymbol{X}_N^{\{t\}})$. The key step for prediction is to obtain the subject $N$'s subject-specific random intercept $b_N^*$ and random function $b_N(s)$. This could be achieved by sampling $b_N^*$ and FPC scores vector $\boldsymbol{\xi}_N$ jointly from their posterior distribution $p(\boldsymbol{b}_N|T_N^* > t, \boldsymbol{y}_N^{\{t\}}, \boldsymbol{\theta})$, where $\boldsymbol{b}_N = [b_N^*, \boldsymbol{\xi}_N]$, and reconstructing random function $b_N(s) = \boldsymbol{\xi}_N\boldsymbol{B}_\phi\boldsymbol{\psi}(s)^\top$. Conditional on the $d$th posterior sample $\boldsymbol{\theta}^{(d)}$, $d = 1, \ldots, D$, we draw the $d$th sample of the $\boldsymbol{b}_N$ from the posterior distribution

$$\begin{aligned} p(\boldsymbol{b}_N|T_N^* > t, \boldsymbol{y}_N^{\{t\}}, \boldsymbol{\theta}^{(d)}) &= \frac{p(\boldsymbol{y}_N^{\{t\}}, T_N^* > t, \boldsymbol{b}_N|\boldsymbol{\theta}^{(d)})}{p(\boldsymbol{y}_N^{\{t\}}, T_N^* > t|\boldsymbol{\theta}^{(d)})} \propto p(\boldsymbol{y}_N^{\{t\}}, T_N^* > t, \boldsymbol{b}_N|\boldsymbol{\theta}^{(d)}) \\ &= p(\boldsymbol{y}_N^{\{t\}}|\boldsymbol{\theta}^{(d)}, \boldsymbol{b}_N)p(T_N^* > t|\boldsymbol{\theta}^{(d)}, \boldsymbol{b}_N)p(\boldsymbol{b}_N|\boldsymbol{\theta}^{(d)}). \end{aligned}$$

where $p(\boldsymbol{y}_N^{\{t\}}|\boldsymbol{\theta}^{(d)}, \boldsymbol{b}_N)$ is the joint conditional probability of longitudinal scalar and functional outcomes, $p(T_N^* > t|\boldsymbol{\theta}^{(d)}, \boldsymbol{b}_N)$ is the survival probability, and $p(\boldsymbol{b}_N|\boldsymbol{\theta}^{(d)})$ is the probability of random effect. For each of $\boldsymbol{\theta}^{(d)}$, $d = 1, \ldots, D$, we use adaptive rejection Metropolis sampling (ARMS) (Gilks et al., 1995) to draw one sample of random effect vector $\boldsymbol{b}_N$. This process is repeated for the $D$ saved values of $\boldsymbol{\theta}$ so that $D$ samples of random effect vector $\boldsymbol{b}_N$ are obtained. The predictions can be calculated by plugging in the samples of the parameter vector and random effect vector $\{\boldsymbol{\theta}^{(d)}, \boldsymbol{b}_N^{(d)}, d = 1, \ldots, D\}$ into the proposed models. For example, based on model (1), the expected values of the longitudinal scalar outcome for subject $N$ at time $t'$ are calculated with respect to the posterior distribution of the parameters $\{\boldsymbol{\theta}|\mathcal{D}_I\}$ as

$$\begin{aligned} &E\{y_N^*(t')|T_N^* > t, \boldsymbol{y}_N^{\{t\}}, \boldsymbol{X}_N^{\{t\}}, \mathcal{D}_I\} \\ &= \int E\{y_N^*(t')|T_N^* > t, \boldsymbol{y}_N^{\{t\}}, \boldsymbol{X}_N^{\{t\}}, \boldsymbol{\theta}\}p(\boldsymbol{\theta}|\mathcal{D}_I)d\boldsymbol{\theta}, \end{aligned} \tag{7}$$

where $\mathcal{D}_I$ denotes the sample on which the model is fitted. The first part of the integrand is given as

$$\begin{aligned} &E\{y_N^*(t')|T_N^* > t, \boldsymbol{y}_N^{\{t\}}, \boldsymbol{X}_N^{\{t\}}, \boldsymbol{\theta}\} \\ &= \int E\{y_N^*(t')|T_N^* > t, \boldsymbol{y}_N^{\{t\}}, \boldsymbol{X}_N^{\{t\}}, b_N^*, \boldsymbol{\theta}\}p(b_N^*|T_N^* > t, \boldsymbol{y}_N^{\{t\}}, \boldsymbol{X}_N^{\{t\}}, \boldsymbol{\theta})db_N^* \\ &= \int \{\beta_0 + t'\beta_t + \sum_p^P x_{Np}\beta_p + \boldsymbol{V}_R(t')\boldsymbol{\zeta} + b_N^*\}p(b_N^*|T_N^* > t, \boldsymbol{y}_N^{\{t\}}, \boldsymbol{X}_N^{\{t\}}, \boldsymbol{\theta})db_N^* \\ &= \beta_0 + t'\beta_t + \sum_p^P x_{Np}\beta_p + \boldsymbol{V}_R(t')\boldsymbol{\zeta} + \int b_N^*p(b_N^*|T_N^* > t, \boldsymbol{y}_N^{\{t\}}, \boldsymbol{X}_N^{\{t\}}, \boldsymbol{\theta})db_N^*. \end{aligned} \tag{8}$$

The integration with respect to $\boldsymbol{\theta}$ in Eq. (7) and the integration with respect to $b_N^*$ in Eq. (8) can be approximated using a Monte Carlo simulation scheme (Rizopoulos, 2011), where the $d$th Monte Carlo sample is

$$E^{(d)}\{y_N^*(t')|T_N^* > t, \boldsymbol{y}_N^{\{t\}}, \boldsymbol{X}_N^{\{t\}}, \mathcal{D}_I\} = \beta_0^{(d)} + t'\beta_t^{(d)} + \sum_p^P x_{Np}\beta_p^{(d)} + \boldsymbol{V}_R(t')\boldsymbol{\zeta}^{(d)} + (b_N^*)^{(d)}, \quad d = 1, \ldots, D.$$

Similarly, based on model (4), the expected values of the longitudinal functional outcome for subject $N$ at time $t'$ is

$$\begin{aligned} &E\{y_N(s, t')|T_N^* > t, \boldsymbol{y}_N^{\{t\}}, \boldsymbol{X}_N^{\{t\}}, \mathcal{D}_I\} \\ &= \int E\{y_N(s, t')|T_N^* > t, \boldsymbol{y}_N^{\{t\}}, \boldsymbol{X}_N^{\{t\}}, \boldsymbol{\theta}\}p(\boldsymbol{\theta}, |\mathcal{D}_I)d\boldsymbol{\theta}, \end{aligned}$$

where the first part of the integrand is given as

$$\begin{aligned} & E\{y_N(s,t')|T_N^* > t, \boldsymbol{y}_N^{\{t\}}, \boldsymbol{X}_N^{\{t\}}, \boldsymbol{\theta}\} \\ = & \int E\{y_N(s,t')|T_N^* > t, \boldsymbol{y}_N^{\{t\}}, \boldsymbol{X}_N^{\{t\}}, \boldsymbol{\xi}_N, \boldsymbol{\theta}\} p(\boldsymbol{\xi}_N|T_N^* > t, \boldsymbol{y}_N^{\{t\}}, \boldsymbol{X}_N^{\{t\}}, \boldsymbol{\theta}) d\boldsymbol{\xi}_N \\ = & \boldsymbol{B}_0\boldsymbol{\psi}^\top(s) + t'\boldsymbol{B}_t\boldsymbol{\psi}^\top(s) + \sum_p^P x_{Np}\boldsymbol{B}_p\boldsymbol{\psi}^\top(s) + \int \boldsymbol{\xi}_N\boldsymbol{B}_\phi\boldsymbol{\psi}^\top(s) p(\boldsymbol{\xi}_N|T_N^* > t, \boldsymbol{y}_N^{\{t\}}, \boldsymbol{X}_N^{\{t\}}, \boldsymbol{\theta}) d\boldsymbol{\xi}_N. \end{aligned}$$

In addition, based on model (5), the conditional probability of event-free at time $t'$ is

$$\begin{aligned} \pi_N(t'|t) = & \int P(T_N^* \geq t'|T_N^* > t, \boldsymbol{y}_N^{\{t\}}, \boldsymbol{X}_N^{\{t\}}, \boldsymbol{\theta}) p(\boldsymbol{\theta}, |\mathcal{D}_l) d\boldsymbol{\theta}, \quad \text{and} \\ P(T_N^* \geq t'|T_N^* > t, \boldsymbol{y}_N^{\{t\}}, \boldsymbol{X}_N^{\{t\}}, \boldsymbol{\theta}) = & \int P(T_N^* \geq t'|T_N^* > t, \boldsymbol{y}_N^{\{t\}}, \boldsymbol{X}_N^{\{t\}}, \boldsymbol{b}_N) p(\boldsymbol{b}_N|T_N^* > t, \boldsymbol{y}_N^{\{t\}}, \boldsymbol{X}_N^{\{t\}}, \boldsymbol{\theta}) d\boldsymbol{b}_N \\ = & \int \frac{P(T_N^* \geq t'|T_N^* > t, \boldsymbol{y}_N^{\{t\}}, \boldsymbol{X}_N^{\{t\}}, \boldsymbol{b}_N)}{P(T_N^* \geq t|T_N^* > t, \boldsymbol{y}_N^{\{t\}}, \boldsymbol{X}_N^{\{t\}}, \boldsymbol{b}_N)} p(\boldsymbol{b}_N|T_N^* > t, \boldsymbol{y}_N^{\{t\}}, \boldsymbol{X}_N^{\{t\}}, \boldsymbol{\theta}) d\boldsymbol{b}_N. \end{aligned}$$

The Monte Carlo samples of $E\{y_N(s,t')|T_N^* > t, \boldsymbol{y}_N^{\{t\}}, \boldsymbol{X}_N^{\{t\}}, \mathcal{D}_l\}$ and $\pi_N(t'|t)$ can be obtained by simply replacing $\{\boldsymbol{\theta}, \boldsymbol{b}_N\}$ in the model with $\{\boldsymbol{\theta}^{(d)}, \boldsymbol{b}_N^{(d)} : d = 1, \ldots, D\}$. All prediction results can then be obtained by calculating simple summaries (e.g., mean, variance, quantiles) of the $D$ samples.

Suppose that subject $N$ has not experienced the event of interest by time $t'$, then the outcome histories are updated to $\boldsymbol{y}_N^{\{t'\}}$. We can dynamically update the posterior distribution to $p(\boldsymbol{b}_N|T_N^* > t', \boldsymbol{y}_N^{\{t'\}}, \boldsymbol{\theta})$, draw new samples, and obtain the updated predictions. We assess the performance of the proposed predictive measures in discriminate between patients who had the event from patients who did not. Such discrimination performance is measured by the integrated area under the time-dependent receiver operating characteristic curve that accommodates censoring time (Li et al., 2016).

## 4. Application to the ADNI study

We apply the proposed Bayesian MFJM to the motivating ADNI-1 study. Besides the longitudinal ADAS-Cog 11 and imaging marker, we include the following variables as scalar covariates: baseline age (*bAge*, mean: 74.4, SD: 7.3, range 55.1–89.3), gender (*gender*, 36.1% female), years of education (*Edu*, mean: 15.6, SD: 3.0, range 4–20), and presence of at least one apolipoprotein E-$\varepsilon$4 allele ($APOE - \varepsilon4$, 56%), given their potential effects on AD progression (Risacher et al., 2009; Cui et al., 2011; Li et al., 2013). To investigate the different forms of imaging information, we include the baseline hippocampal volume (*bHV*), baseline hippocampal surface based on hippocampal radial distance (*bHRD*), and longitudinal hippocampal radial distance (*lHRD*).

We proposed three joint models with the same longitudinal submodel of the scalar outcome, ADAS-Cog 11, which is defined by

$$\begin{aligned} ADAS - Cog_i(t_{ij}) = & \ m_i(t_{ij}) + \varepsilon_{ij} \\ m_i^*(t_{ij}) = & \ \beta_0 + \beta_1 APOE - \varepsilon4_i + \beta_2 bAge_i + \beta_t t_{ij} + \sum_{r=1}^3 \zeta_r (t_{ij} - \kappa_r)_+ + b_{i1}^*. \end{aligned}$$

The three joint models are varied in the survival part that incorporate different levels of imaging information in prediction of AD progression. In the first joint model (refer to as model *JM*), we incorporate the baseline hippocampal volume (*bHV*) as a scalar predictor, along with other covariates and underline process $m_i^*(t)$ of the ADAS-Cog 11, in the survival part. This gives the survival submodel in *JM* as

$$h_i(t) = h_0(t) \exp\{\gamma_1 gender_i + \gamma_2 bAge_i + \gamma_3 Edu_i + \gamma_4 APOE - \varepsilon4 + \gamma_5 bHV_i + \alpha^* m_i^*(t)\}.$$

The second model is a function joint model (refer to as model *FJM*) which includes baseline hippocampal radial distance *bHRD*$(s)$, instead of hippocampal volume, as a time-independent functional predictor in the survival submodel. The model was proposed and applied to ADNI study in the previous work (Li and Luo, 2017a), in which the survival submodel is defined as

$$\begin{aligned} h_i(t) = & \ h_0(t) \exp\{\gamma_1 gender_i + \gamma_2 bAge_i + \gamma_3 Edu_i + \gamma_4 APOE - \varepsilon4 \\ & + \int_S bHRD_i(s) B_{bHRD}(s) ds + \alpha^* m_i^*(t)\}. \end{aligned}$$

The third model is a multivariate functional joint model (refer to as model *MFJM*) that accounts for the longitudinal hippocampal radial distance *lHRD*$(s,t)$ in the survival submodel, where *lHRD*$(s,t)$ is modeled as

$$\begin{aligned} lHRD_i(s,t_{ij}) = & \ m_i(s,t_{ij}) + \varepsilon_{ij}(s) \\ m_i(s,t_{ij}) = & \ B_0(s) + B_1(s) APOE - \varepsilon4_i + B_2(s) bAge_i + B_t(s) t_{ij} + b_{i1}(s), \end{aligned}$$

**Table 1**
Areas under the ROC curve (AUC) by three candidate models in the ADNI study.

| $\Delta t$ | $t$ | JM 1 | FJM | MFJM |
|---|---|---|---|---|
| 0.5 | 1 | 0.715 | 0.754 | 0.821 |
| | 1.5 | 0.691 | 0.738 | 0.734 |
| | 2 | 0.781 | 0.809 | 0.812 |
| 1 | 1 | 0.696 | 0.747 | 0.792 |
| | 1.5 | 0.735 | 0.776 | 0.777 |
| | 2 | 0.749 | 0.769 | 0.766 |

and the survival submodel is

$$h_i(t) = h_0(t)\exp\{\gamma_1 gender_i + \gamma_2 Edu_i + +\gamma_3 bAge_i + \gamma_4 APOE - \varepsilon 4_i + \alpha^* m_i^*(t) + \alpha \int_{\boldsymbol{S}} m_i(s,t)ds\}.$$

In the *MFJM*, we expand the random functions $b_{i1}(s) = \sum_{k=1}^{K_{\boldsymbol{\phi}}=4} \xi_{ik}\phi_k(s)$, and consider the correlation between $b_{i1}^*$ and $\xi_{i1}$. We express coefficient functions $B_p(s)$ and eigenfunctions $\phi_k(s)$ in terms of a known cubic *B*-spline basis functions $\boldsymbol{\psi}(s)$ with 10 knots. We allow a flexible and smooth disease progression along time by using truncated power series splines with 3 knots at the location $\boldsymbol{\kappa} = (1, 2, 3)$ in years, which ensures sufficient patients within each interval. Baseline hazard function $h_0(t)$ is approximated by a piecewise constant function. Specifically, the observed survival time is divided into $H = 7$ intervals by every $1/H$th quantiles. We have also explored other selections of $K_{\boldsymbol{\phi}}$ and $H$ and obtained very similar results.

The three candidate models are compared via assessing their predictive performance, manifested by the time-dependent AUCs, at different time points over the follow-up period. To avoid overestimation of the prediction, we conduct a 10-fold cross validation. Parameters of the joint model are estimated from the training dataset and applied to the validation dataset. The conditional event-free probability corresponding to the time frame $(t, t + \Delta t]$ is predicted for each patient in the validation datasets as described in Section 3.3. Because the ADNI patients were reassessed approximately every half year, we select $t$ at 1, 1.5, and 2 years, and $\Delta t$ as 0.5 and 1 years for analysis. Then the time-dependent AUCs are calculated based on the predicted probabilities of all patients.

Table 1 displays the time-dependent AUCs from the three candidate models. Models *FJM* and *MFJM* have notably larger AUCs than model *JM* for most combinations of $t$ and $\Delta t$. This suggests that including functional predictor *HRD* in the survival submodel improves the capability of the joint model in predicting risk of AD diagnosis. However, the predictive capability of longitudinal *HRD* does not show much advantage than the baseline *HRD* information, except the early phase of the follow-up. This may be explained by the fact that different markers may be more or less discriminative at different stages of disease, and MRI abnormalities usually occurring earlier before any symptom of cognitive impairment appears (Jack Jr et al., 2010; Li et al., 2013). We have also assessed the *MFJM* accounting for the correlations between $b_{i1}^*$ and the first two FPC components $[\xi_{i1}, \xi_{i2}]$. The model has no notable improvement in terms of prediction.

We select *MFJM* as the final model because it has a competitive good discrimination capability in both early and late phases. Parameter estimates from model *MFJM* using whole dataset are presented in Table 2. In the longitudinal submodel of scalar outcome, people with *APOE*-$\varepsilon$4 allele(s), on average, have higher (worse) ADAS-Cog 11 score (2.207 unit) than people without this genetic variation. Also, the ADAS-Cog 11 score increases (deteriorates) as time progresses, i.e., an average increase of 1.121 unit (95% CI: [0.648–01.584]) per year for MCI patients. In the survival submodel, the presence of *APOE*-$\varepsilon$4 allele(s) increases the hazard of AD diagnosis by 51% ($\exp(0.409) - 1$, 95% CI: [8%-109%]), which is consistent with the literature (Corder et al., 1993). Furthermore, larger ADAS-Cog 11 score increases the risk of AD diagnosis, i.e., one unit increase in ADAS-Cog 11 score increases the hazard of AD diagnosis by 19% ($\exp(0.173) - 1$, 95% CI: [14%–24%]). The association parameter $\alpha$ is negative, indicating that the decrease of HRD (i.e., hippocampal atrophy) is associated with the increasing risk of AD diagnosis. The estimated coefficient functions in the longitudinal model for functional outcome HRD is presented in Fig. 2. *APOE*-$\varepsilon$4 allele(s) is not associated with the thickness of hippocampus because the estimated coefficient function $\hat{B}_1(s)$ (upper right panel) fluctuates around zero across the domain $S$. The estimated coefficient function $\hat{B}_t(s)$ (lower left panel) quantifies the change of HRD over time, and notable atrophy can be viewed on both ends and middle part of hippocampus. The baseline age of the patients has a similar effect on hippocampus. As shown in the estimated coefficient function $\hat{B}_2(s)$ (lower right panel), the older patients, on average, have thinner hippocampus on both ends and middle parts.

To illustrate the personalized dynamic predictions, we select two target patients as validation data, and predict their future health outcomes and event-free probabilities based on *MFJM* estimated using the remaining data as training set. Patient A had a baseline age of 73, no *APOE*-$\varepsilon$4, as compared with a more severe Patient B, 78 years old at baseline, and with *APOE*-$\varepsilon$4. Fig. 3 demonstrates how the predicted ADAS-Cog 11 scores are updated over time for the two patients. From the left to the right on Fig. 3, by using more follow-up data, predictions are closer to the true observed values and the 95% uncertainty band is narrower. Fig. 4 shows the predicted *HRD* at third and fourth visits based on the previous observations. The predicted expected *HRD* (red line) is close to the true observation (black line), however we do not observe significant difference between HRD at the third and the fourth visits and we suggest to use the predicted *HRD* with caution. Fig. 5

**Table 2**
ADNI data analysis results from model *MFJM*.

| | Parameters | Mean | SE | 2.5% | 97.5% |
|---|---|---|---|---|---|
| For scalar longitudinal outcome | | | | | |
| ADAS-Cog 11 | *APOE-ε*4 | 2.207 | 0.397 | 1.464 | 3.006 |
| | *bAge* | 0.172 | 0.258 | −0.346 | 0.668 |
| | Time (Years) | 1.121 | 0.240 | 0.648 | 1.584 |
| For survival process | | | | | |
| MCI to AD | Female | 0.094 | 0.172 | −0.241 | 0.442 |
| | *bAge* | −0.113 | 0.086 | −0.278 | 0.059 |
| | *Edu* (years) | 0.029 | 0.026 | −0.022 | 0.080 |
| | *APOE-ε*4 | 0.409 | 0.164 | 0.081 | 0.736 |
| | $\alpha^*$ | 0.173 | 0.021 | 0.135 | 0.215 |
| | $\alpha$ | −1.105 | 0.437 | −1.969 | −0.259 |

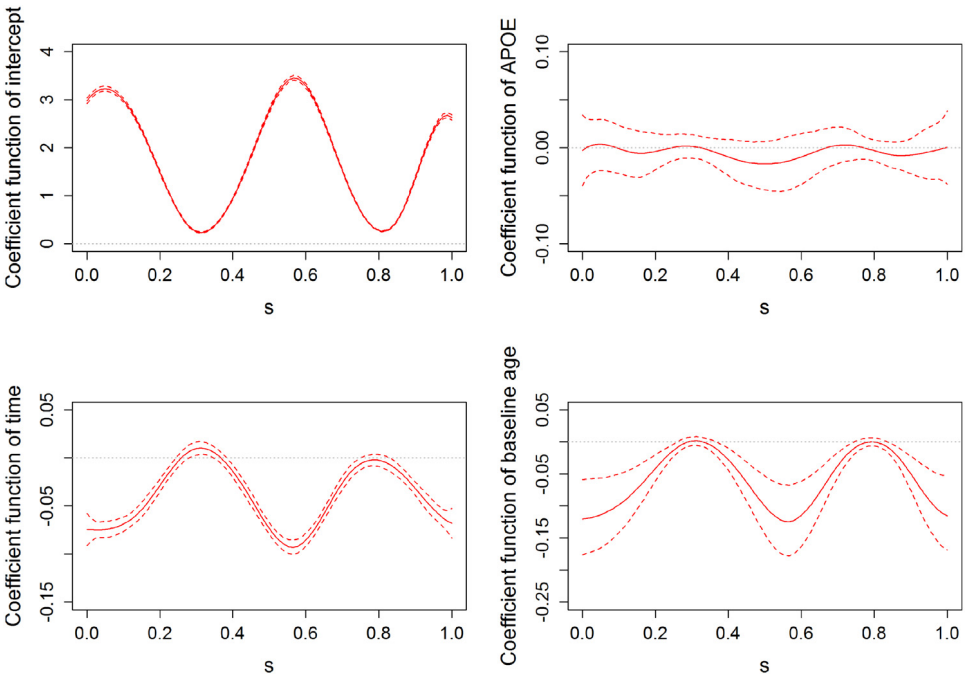


**Fig. 2.** Estimated coefficient functions (solid lines) in the functional longitudinal submodel with 95% pointwise uncertainty band (dashed lines) and reference lines (dotted lines) at zero.

displays the predicted probability of being free of AD diagnosis. For Patient A, the event-free probability curve does not show large change because Patient A's predicted ADAS-Cog 11 scores are relatively low. In comparison, Patient B has higher predicted ADAS-Cog 11 scores and worse cognitive function, and thus has considerable drop in the event-free probability. This suggests that Patient B has a higher risk of AD diagnosis and should be monitored frequently.

## 5. Simulation study

In this section, we conduct a simulation study to evaluate the proposed models. We generate 100 datasets with sample size $I = 150$ subjects and each subject has $J_i = 4$ measurements at time 0, 5, 10, and 15. The simulated data structure is similar to the motivating ADNI study. We generate longitudinal functional response $y_i(s, t_{ij})$ on an equally spaced grid of length 25 and longitudinal scalar response $y_i^*(t_{ij})$ according to the longitudinal submodels

$$\begin{aligned} y_i^*(t_{ij}) &= m_i^*(t_{ij}) + \epsilon_{ij}^*, \\ y_i(s, t_{ij}) &= m_i(s, t_{ij}) + \epsilon_{ij}(s), \text{ where} \\ m_i^*(t_{ij}) &= \beta_0 + t_{ij}\beta_t + x_{i1}\beta_1 + b_i^*, \text{ and} \\ m_i(s, t_{ij}) &= B_0(s) + t_{ij}B_t(s) + x_{i1}B_1(s) + \sum_{k=1}^{K_\phi=2} \xi_{ik}\phi_k(s), \quad s \in [0, 1] = \boldsymbol{S}. \end{aligned}$$

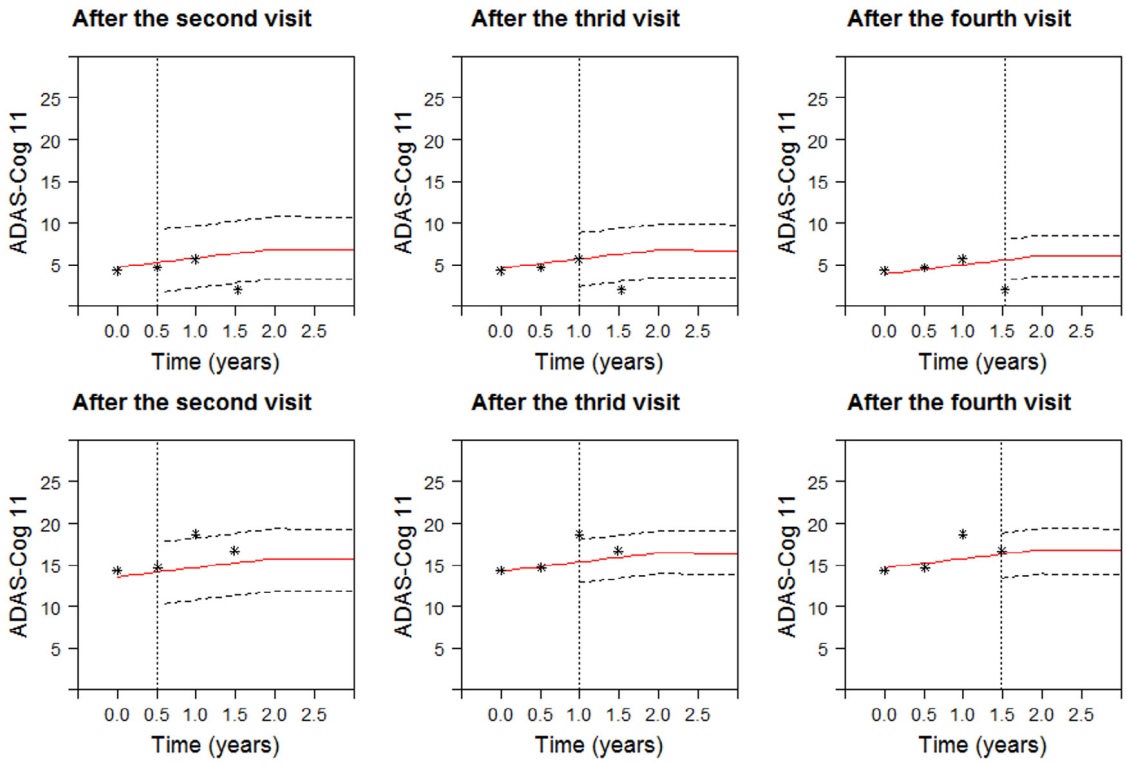


**Fig. 3.** Predicted ADAS-Cog 11 for Patient A (upper panels) and Patient B (lower panels). Solid line is predicted longitudinal trajectories. Dashed lines construct a 95% pointwise uncertainty band. The dotted vertical lines represent the time of prediction $t$.

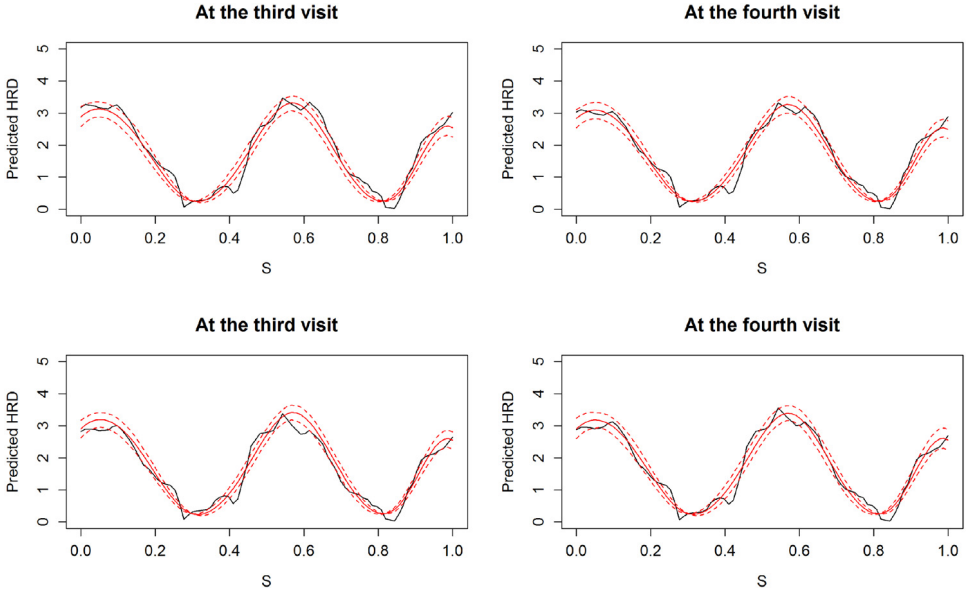


**Fig. 4.** Predicted *HRD* (red) with 95% pointwise uncertainty band (dashed lines) for Patient A (upper panels) and Patient B (lower panels) at third and forth visit. The solid black curve represents the true observation at the time point. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

We set $\beta_0 = 4$, $\beta_t = 1$, and $\beta_1 = 0.5$. The intercept function is $B_0(s) = -1.5 - \sin(2\pi s) - \cos(2\pi s)$. The time effect is $B_t(s) = \frac{1}{10}\Phi(\frac{s-0.5}{0.2^2})$, where $\Phi(.)$ is the standard Normal density function. The fixed effect is $B_1(s) = \sin(2\pi s) - \cos(4\pi s)$, and we generate scalar predictors using $x_{i1} \sim N(0, 2)$. Similar to previous simulation study (Goldsmith and Kitago, 2016),

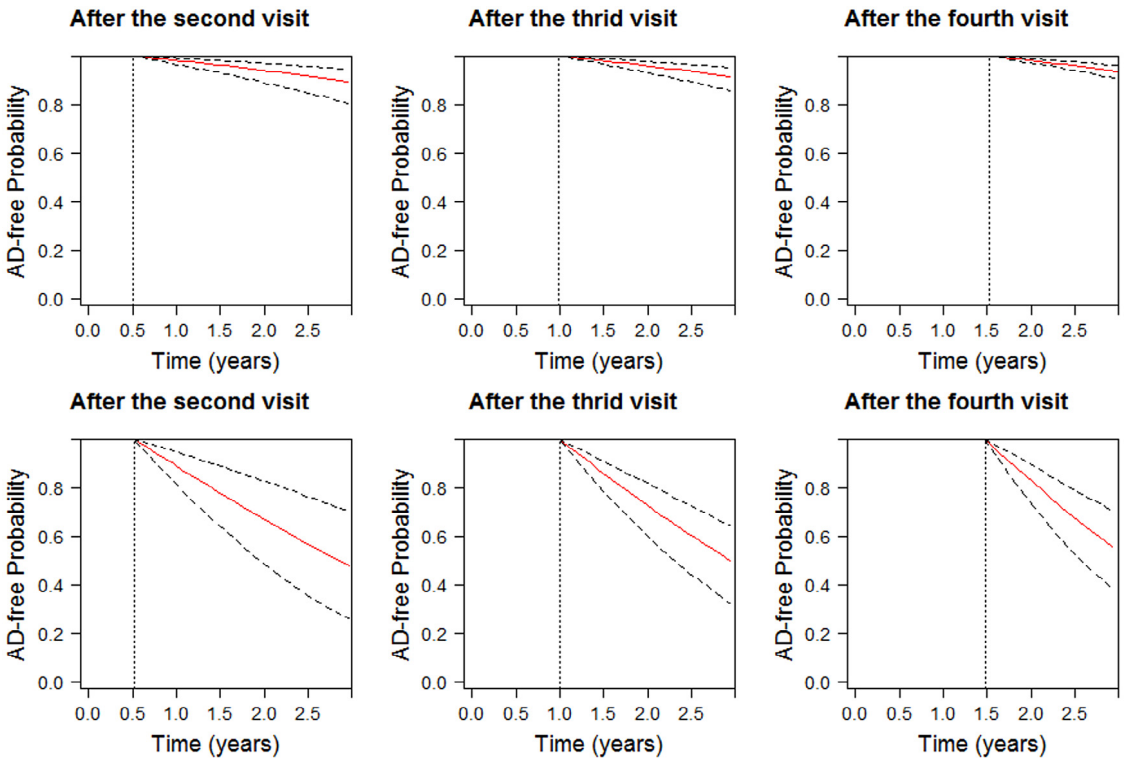


**Fig. 5.** Predicted event-free probability with 95% pointwise uncertainty band (dashed lines) for Patient A (upper panels) and Patient B (lower panels). The dotted vertical lines represent the time of prediction $t$.

the orthogonal eigenfunctions for random intercept functions are chosen to be $\phi_1(s) \propto 1.5 - \sin(2\pi s) - \cos(2\pi s)$ and $\phi_2(s) \propto \sin(4\pi s)$ and scaled such that $\int_{\mathcal{S}}[\phi_1(s)]^2 ds = 1$ and $\int_{\mathcal{S}}[\phi_2(s)]^2 ds = 1$. The subject-specific random effect and FPC scores $[b_i^*, \xi_{i1}, \xi_{i2}]$ are generated from multivariate normal distribution with zero-mean and covariance matrix $\Sigma$ with $\sigma^2_{b^*} = 0.6^2$, $\lambda_1 = 1$, $\lambda_2 = 0.5$ and $\rho = 0.5$, respectively. The measurement error for scalar response $\epsilon_{ij}$ is simulated from $N(0, 0.2)$. The white noise component $\epsilon_{ij}(s)$ is simulated from $N(0, 0.1)$ across $s$.

We choose a constant baseline hazard function $h_0(t) = 0.01$ and the survival submodel is

$$h_i(t) = h_0(t)\exp\{w_i\gamma_1 + \alpha^* m_i^*(t) + \alpha \int_{\mathcal{S}} m_i(s,t)ds\},$$

where $w_i$ is simulated from $N(0, 1)$, $\gamma_1 = 0.76$, $\alpha^* = 0.5$, and $\alpha = 0.3$. We generated random survival times based on the closed-form of $T^*$ derived from survival function:

$$S_i(t) = \exp\Big\{-\frac{\lambda \exp\big(w_i\gamma_1 + \alpha^*(\beta_0 + \beta_1 x_{i1} + b_i^*) + \alpha \int_{\mathcal{S}}[B_0(s) + x_{i1}B_1(s) + \sum_{k=1}^2 \xi_{ik}\phi_k(s)]ds\big)}{\alpha^*\beta_t + \alpha \int_{\mathcal{S}} B_t(s)ds} \times$$
$$[\exp\big(t \times (\alpha^*\beta_t + \alpha \int_{\mathcal{S}} B_t(s)ds)\big) - 1]\Big\},$$

and thus,

$$T_i^* = \log\Big\{\frac{-\log(S_i(t)) \times (\alpha^*\beta_t + \alpha \int_{\mathcal{S}} B_t(s)ds)}{h_0(t)\exp\big(w_i\gamma_1 + \alpha^*(\beta_0 + \beta_1 x_{i1} + b_i^*) + \alpha \int_{\mathcal{S}}[B_0(s) + x_{i1}B_1(s) + \sum_{k=1}^2 \xi_{ik}\phi_k(s)]ds\big) + 1}\Big\} /$$
$$(\alpha^*\beta_t + \alpha \int_{\mathcal{S}} B_t(s)ds),$$

where $S_i(t)$ is simulated from a uniform distribution between 0 and 1. Censoring time is independently simulated from another uniform distribution to achieve a censoring rate about 30%. Due to censoring, each subject has an average of 3 repeated measurements.

For estimation, the coefficient functions $B_0(s)$, $B_t(s)$ , $B_1(s)$, and FPC eigenfunctions $\phi_k(s)$, $k = 1, \ldots, K_\phi$ are expanded by cubic $B$-spline basis with $K_\psi = 10$. We set the number of estimated principal components $\widehat{K}_\phi \in \{2, 3\}$. Note that when

**Table 3**
Parameter estimation in the simulation study based on 100 datasets when $\widehat{K}_{\phi} = 2$.

| | Bias | AMSE | SE | SD | CP |
|---|---|---|---|---|---|
| $\beta_0 = 4$ | <0.001 | 0.003 | 0.053 | 0.055 | 0.930 |
| $\beta_t = 1$ | −0.003 | <0.001 | 0.019 | 0.019 | 0.960 |
| $\beta_1 = 0.5$ | <0.001 | <0.001 | 0.004 | 0.003 | 0.980 |
| $\gamma_1 = 0.76$ | −0.017 | 0.017 | 0.116 | 0.122 | 0.920 |
| $\alpha^* = 0.5$ | 0.031 | 0.005 | 0.059 | 0.062 | 0.910 |
| $\alpha = 0.3$ | −0.023 | 0.053 | 0.232 | 0.231 | 0.930 |

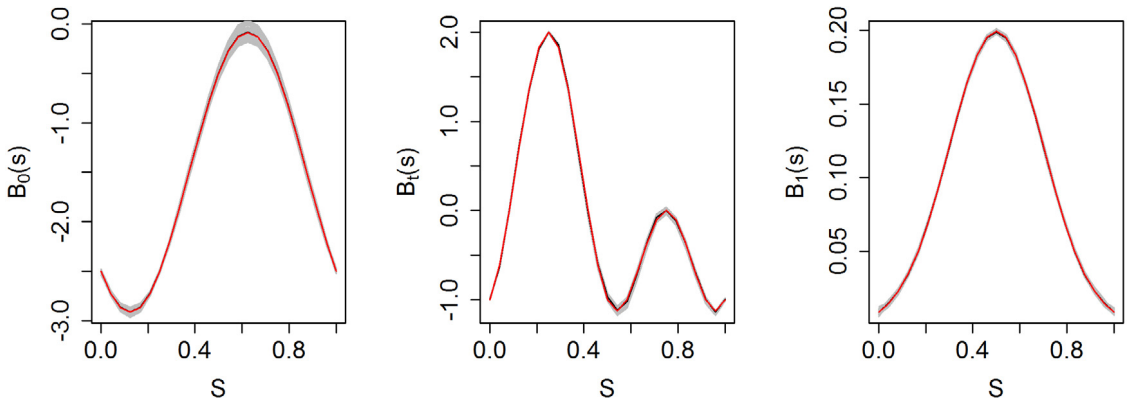


**Fig. 6.** The estimates of the coefficient functions in the simulation study based on 100 runs and $\widehat{K}_{\phi} = 2$. The red solid lines are mean estimated curves and the gray solid lines are 100 estimated curves based on each individual dataset. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

$\widehat{K}_{\phi} = 3$ the number of estimated FPC components is larger than the number of true FPC components, which is held at $K_{\phi} = 2$. Model parameters are estimated for the $C = 100$ datasets using the methodology described in Section 3.2. Estimation and inference are based on posterior means and quartiles of 5000 iteration from the sampler after discarding the first 1000 as burn-in. We perform diagnostics for one simulated dataset indicate that these levels are sufficient for convergence and exploration of the full posterior distribution. On average, the computing process takes 1.1 h for each simulated dataset on a personal computer (RAM 8G, CPU 3.30 GHz). The ability to estimate the true coefficient is assessed by the average mean squared error (AMSE), e.g., $AMSE(\widehat{\beta}_1) = \frac{1}{100}\sum_{c=1}^{100}(\widehat{\beta}_1 - \beta_1)^2$. Table 3 presents the AMSE, in addition to bias (the average of the posterior means minus the true values), standard error (SE, the square root of the average of the variance), standard deviation (SD, the standard deviation of the posterior mean), and coverage probabilities (CP) of 95% credible intervals when $\widehat{K}_{\phi} = 2$. Table 3 suggests that the proposed model performs reasonably well with relatively small bias and AMSE values.

Fig. 6 displays the coefficient functions estimated under $\widehat{K}_{\phi} = 2$. The true coefficient functions (black solid lines), their mean estimated curves (red solid lines), along with 100 estimated curves based on each individual dataset (gray solid lines). The figures suggest that the estimated coefficient functions from the model are reasonably close to the true coefficient functions. The simulation results for scenario $\widehat{K}_{\phi} = 3$ are presented in Web Table 1 and Web Figure 2. It suggests that increasing $\widehat{K}_{\phi}$ has limited effects on parameter estimation in either longitudinal sub-models and survival sub-models. The bias and AMSE values have slight increase but are still relatively small. The estimated coefficient functions are reasonably close to the true coefficient functions.

For each testing dataset, we predict subject-specific conditional survival probability at different time points $t$ and $\Delta t$ using the MCMC samples from the fitted model and available measurements up to time $t$. Table 4 presents the time-dependent AUCs by averaging the separate analyses of 100 datasets. The true AUCs are computed using the prespecified parameter values and random effects when generating the data. The predicted AUCs are relatively close to the true AUCs, suggesting a good prediction performance of the MFJM in terms of validation.

## 6. Discussion

The proposed multivariate functional joint model (MFJM) is an important complement to both functional data analysis and joint modeling for longitudinal and survival data. Our model allows the longitudinal functional outcome, longitudinal scalar outcome, and survival outcome to be modeled simultaneously and can be applied to many areas of research when MRI data and clinical variables are collected longitudinally. We use the functional mixed effect model and functional principal component (FPC) analysis techniques to approximate the longitudinal functional data, and expand the coefficient functions

**Table 4**
Areas under the ROC curve (AUC) for the simulation study.

| $\Delta t$ | $t$ | True | $AUC_t^{\Delta t}$ |
|---|---|---|---|
| 5 | 0 | 0.831 | 0.830 |
| | 5 | 0.811 | 0.810 |
| | 10 | 0.829 | 0.827 |
| 10 | 0 | 0.847 | 0.846 |
| | 5 | 0.875 | 0.875 |
| | 10 | 0.824 | 0.823 |

and eigenfunctions in the model using a penalized spline approach. We then develop the process of making personalized dynamic prediction of future outcomes and risks of event of interest using both repeated functional and scalar outcomes.

The model inference is conducted using a Bayesian approach. One advantage of Bayesian approach is the availability of Markov chain Monte Carlo (MCMC) sampling algorithms, which allow estimation from posterior probability density functions that are not analytically tractable, and which require complex multi-dimensional integration over the random effects. The surge in MCMC sampling can be partly explained by the wide use of the Bayesian computing languages, such as `Stan`, which eliminate the need of complex analytical derivation of the posterior distributions. Moreover, Bayesian method is well-suited for dynamic prediction using joint models. With the MCMC samples from the posterior distributions of the parameters for the original data, we can devise a simple simulation scheme to obtain a Monte Carlo estimate of risk prediction and longitudinal trajectories (Rizopoulos, 2011; Rizopoulos et al., 2014). In addition, uncertainty about posterior parameter estimates is readily calculated from the MCMC output without the need for further complex derivation and calculations. This facilitates the calculation of the uncertainty intervals of functional coefficients and functional predictions.

Simulation indicates that the proposed Bayesian MFJM yields accurate inference and prediction. The application of our developed methodology to the motivating data yields novel insight into the effect of regional atrophy on the AD progression. More importantly, the proposed dynamic prediction approach can utilize the functional and scalar predictors to make correct predictions for new subjects. The inclusion of longitudinal functional predictor hippocampal radial distance (HRD) into the survival submodel improves the predictive performance in the early phase of disease among MCI patients. When new measurements are available, the predictions can be updated with improved accuracy and efficiency. The practical impact of such dynamic prediction tools can be dramatic for the neurodegenerative diseases (e.g., Alzheimer's disease) because the longitudinal functional data are increasingly collected in the studies of these diseases. They provide unique insight and valuable guidance for clinical decision making on patient prognosis, targeted treatment, and for targeted recruitment for clinical trials.

There are some limitations we will address in the future. First, our method requires to select the number of the FPC components prior to analysis. Although we suggest a large number of FPC components, determining whether a selection is sufficient to describe the major variation in the functional data require additional analyses with even larger values, which can incur considerable computational expense. Second, we jointly model all parameters of interest in a Bayesian context, the computation time of the proposed Bayesian procedure could be a serious concern particularly as sample sizes, dimension of functional data, and the number of estimated principal components grow. Future work focusing on variation Bayes or other approximation may address the computational concern. Another option is to use a two-step method which models longitudinal functional outcome using functional principal components approaches and then plugs in the estimates of FPC scores into the survival model. These approaches are more computationally attractive and scalable than joint modeling approach but may be accompanied by poorer inferential performance (Crainiceanu et al., 2009; Goldsmith et al., 2015). It is worthwhile to investigate when using these methods is a reasonable alternative to the joint analysis. For example, the two-step method may provide tools for choosing the dimension of the FPC components via rapid comparisons of different selections. Third, we assume a constant associate function $\alpha(s) \equiv \alpha$ to quantify the association between the functional outcome and the hazard. If the primary interest is prediction, we can allow $\alpha(s)$ varies over the domain of $\boldsymbol{S}$ and expand it by the cubic $B$-spline basis functions and estimate the spline coefficients. Another direct extension of the model is the inclusion of covariate-specific random effects, such as random time-slope function, in the functional mixed effect model (2). In this paper, we introduce the correlation between the scalar and functional outcomes using the correlation between the scalar random effects and the first FPC component derived from the random function. Accounting for the correlations between the random effects and the first few FPC components may represent the correlation between the scalar and functional outcomes more accurate, but may also lead to increased computing burden. It is worthwhile to further explore other correlation structures, e.g., adopting an idea of latent trial model (Dunson, 2003; Wang et al., 2017), especially when additional covariate-specific random effects are in the model. Moreover, the inverse-gamma prior distribution, which we used for variance parameters in the model inference, can be sensitive to the choice of the hyperparameters (shape and scale) in case where variance estimates are close to zero (Gelman, 2006). We tested the model inference on one training dataset from our application study using uniform and half-Cauchy prior distribution instead, and achieved reasonably similar results. We suggest that further tests are needed in other applications and datasets on a case-by-case basis. We would like to investigate the effect of these extensions and address the limitations to improve predict performance in our future research.

## Acknowledgments

Sheng Luo 's research was supported by the National Institute of Neurological Disorders and Stroke under Award Number R01NS091307. The authors acknowledge the Texas Advanced Computing Center (TACC) for providing high-performing computing resources. Data used in preparation of this article were obtained from Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.ucla.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this article. A complete listing of ADNI investigators can be found at: http://adni.loni.ucla.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf.

## Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.csda.2018.07.015.

## References

Apostolova, L.G., Mosconi, L., Thompson, P.M., Green, A.E., Hwang, K.S., Ramirez, A., Mistur, R., Tsui, W.H., de Leon, M.J., 2010. Subregional hippocampal atrophy predicts alzheimer's dementia in the cognitively normal. Neurobiol. Aging 31 (7), 1077–1088.

Brumback, B.A., Rice, J.A., 1998. Smoothing spline models for the analysis of nested and crossed samples of curves. J. Amer. Statist. Assoc. 93 (443), 961–976.

Chen, K.H.M., Chuah, L.Y.M., Sim, S.K.Y., Chee, M.W.L., 2010. Hippocampal region-specific contributions to memory performance in normal elderly. Brain Cogn. 72 (3), 400–407.

Corder, E., Saunders, A., Strittmatter, W., Schmechel, D., Gaskell, P., Small, G., Roses, A.D., Haines, J., Pericak-Vance, M.A., 1993. Gene dose of apolipoprotein E type 4 allele and the risk of alzheimer's disease in late onset families. Science 261 (5123), 921–923.

Crainiceanu, C.M., Goldsmith, A.J., 2010. Bayesian functional data analysis using WinBUGS. J. Stat. Softw. 32 (11).

Crainiceanu, C.M., Staicu, A.-M., Di, C.-Z., 2009. Generalized multilevel functional regression. J. Amer. Statist. Assoc. 104 (488), 1550–1561.

Cui, Y., Liu, B., Luo, S., Zhen, X., Fan, M., Liu, T., Zhu, W., Park, M., Jiang, T., Jin, J.S., et al., 2011. Identification of conversion from mild cognitive impairment to alzheimer's disease using multivariate predictors. PLoS One 6 (7), e21896.

Di, C.-Z., Crainiceanu, C.M., Caffo, B.S., Punjabi, N.M., 2009. Multilevel functional principal component analysis. Ann. Appl. Stat. 3 (1), 458–488.

Du, A.T., 2001. Magnetic resonance imaging of the entorhinal cortex and hippocampus in mild cognitive impairment and alzheimer's disease. J. Neurol. Neurosurg. Psychiatry 71 (4), 441–447.

Dunson, D.B., 2003. Dynamic latent trait models for multidimensional longitudinal data. J. Amer. Statist. Assoc. 98 (463), 555–563.

Faucett, C.L., Thomas, D.C., 1996. Simultaneously modelling censored survival data and repeatedly measured covariates: A Gibbs sampling approach. Stat. Med. 15 (15), 1663–1685.

Frisoni, G.B., Fox, N.C., Jack, C.R., Scheltens, P., Thompson, P.M., 2010. The clinical use of structural mri in alzheimer disease. Nature Reviews Neurology 6 (2), 67–77.

Gelman, A., 2006. Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper). Bayesian Anal. 1 (3), 515–534.

Gelman, A., Carlin, J.B., Stern, H.S., Rubin, D.B., 2013. Bayesian Data Analysis. CRC Press.

Gilks, W.R., Best, N.G., Tan, K.K.C., 1995. Adaptive rejection Metropolis sampling within Gibbs sampling. J. R. Stat. Soc. Ser. C. Appl. Stat. 44 (4), 455–472.

Goldsmith, J., Kitago, T., 2016. Assessing systematic effects of stroke on motor control by using hierarchical function-on-scalar regression. J. R. Stat. Soc. Ser. C. Appl. Stat. 65 (2), 215–236.

Goldsmith, J., Zipunnikov, V., Schrack, J., 2015. Generalized multilevel function-on-scalar regression and principal component analysis. Biometrics 71 (2), 344–353.

Gomar, J.J., Bobes-Bascaran, M.T., Conejero-Goldberg, C., Davies, P., Goldberg, T.E., Alzheimer's Disease Neuroimaging Initiative, 2011. Utility of combinations of biomarkers, cognitive markers, and risk factors to predict conversion from mild cognitive impairment to alzheimer disease in patients in the alzheimer's disease neuroimaging initiative. Arch. Gen. Psychiatry 68 (9), 961–969.

Greven, S., Crainiceanu, C., Caffo, B., Reich, D., 2010. Longitudinal functional principal component analysis. Electron. J. Stat. 4, 1022–1054.

Guo, W., 2002. Functional mixed effects models. Biometrics 58 (1), 121–128. http://dx.doi.org/10.1111/j.0006-341X.2002.00121.x.

Henderson, R., Diggle, P., Dobson, A., 2000. Joint modelling of longitudinal measurements and event time data. Biostatistics 1 (4), 465–480.

Hickey, G.L., Philipson, P., Jorgensen, A., Kolamunnage-Dona, R., 2016. Joint modelling of time-to-event and multivariate longitudinal outcomes: recent developments and issues. BMC Med. Res. Methodol. 16, 117.

Hoffman, M.D., Gelman, A., 2014. The no-u-turn sampler: adaptively setting path lengths in hamiltonian monte carlo. J. Mach. Learn. Res. 15 (1), 1593–1623.

Huang, M., Yang, W., Feng, Q., Chen, W., Initiative, A.D.N., et al., 2017. Longitudinal measurement and hierarchical classification framework for the prediction of alzheimers disease. Sci. Rep. 7.

Jack Jr., C.R., Knopman, D.S., Jagust, W.J., Shaw, L.M., Aisen, P.S., Weiner, M.W., Petersen, R.C., Trojanowski, J.Q., 2010. Hypothetical model of dynamic biomarkers of the Alzheimer's pathological cascade. Lancet Neurol. 9 (1), 119–128.

Jenkinson, M., Beckmann, C.F., Behrens, T.E., Woolrich, M.W., Smith, S.M., 2012. Fsl. NeuroImage 62 (2), 782–790.

Lang, S., Brezger, A., 2004. Bayesian P-splines. J. Comput. Graph. Statist. 13 (1), 183–212.

Li, K., Chan, W., Doody, R.S., Quinn, J., Luo, S., 2017. Prediction of conversion to Alzheimers disease with longitudinal measures and time-to-event data. J. Alzheimer's Dis. 58 (2), 361–371. http://dx.doi.org/10.3233/JAD-161201.

Li, L., Greene, T., Hu, B., 2016. A simple method to estimate the time-dependent receiver operating characteristic curve and the area under the curve with right censored data. Stat. Methods Med. Res. OnlineFirst. http://dx.doi.org/10.1177/0962280216680239.

Li, K., Luo, S., 2017a. Dynamic predictions in Bayesian functional joint models for longitudinal and time-to-event data: An application to Alzheimers disease. Stat. Methods Med. Res. OnlineFirst.

Li, K., Luo, S., 2017b. Functional joint model for longitudinal and time-to-event data: an application to alzheimer's disease. Stat. Med. 36 (22), 3560–3572.

Li, S., Okonkwo, O., Albert, M., Wang, M.C., 2013. Variation in variables that predict progression from mci to ad dementia over duration of follow-up. Amer. J. Alzheimer's Dis. 2 (1), 12–28.

Liu, F., Li, Q., 2016. A Bayesian model for joint analysis of multivariate repeated measures and time to event data in crossover trials. Stat. Methods Med. Res. 25 (5), 2180–2192.

Morris, J.S., 2015. Functional regression. Annu. Rev. Statist. Appl. 2 (1), 321–359.

Morris, J.S., Carroll, R.J., 2006. Wavelet-based functional mixed models. J. R. Stat. Soc. Ser. B Stat. Methodol. 68 (2), 179–199.

Neal, R.M., et al., 2011. Mcmc using hamiltonian dynamics. Handb. Markov Chain Monte Carlo 2, 113–162.
Park, S.Y., Staicu, A.-M., 2015. Longitudinal functional data analysis. Stat 4 (1), 212–226.
Patenaude, B., Smith, S.M., Kennedy, D.N., Jenkinson, M., 2011. A Bayesian model of shape and appearance for subcortical brain segmentation. NeuroImage 56 (3), 907–922.
Perrin, R.J., Fagan, A.M., Holtzman, D.M., 2009. Multi-modal techniques for diagnosis and prognosis of Alzheimers disease. Nature 461 (7266), 916–922.
Petersen, R.C., Smith, G.E., Waring, S.C., Ivnik, R.J., Tangalos, E.G., Kokmen, E., 1999. Mild cognitive impairment: clinical characterization and outcome. Arch. Neurol. 56 (3), 303–308.
Qiu, A., Fennema-Notestine, C., Dale, A.M., Miller, M.I., Initiative, A.D.N., et al., 2009. Regional shape abnormalities in mild cognitive impairment and alzheimer's disease. NeuroImage 45 (3), 656–661.
Risacher, S.L., Saykin, A.J., Wes, J.D., Shen, L., Firpi, H.A., McDonald, B.C., 2009. Baseline mri predictors of conversion from mci to probable ad in the adni cohort. Curr. Alzheimer Res. 6 (4), 347–361.
Rizopoulos, D., 2011. Dynamic predictions and prospective accuracy in joint models for longitudinal and time-to-event data. Biometrics 67 (3), 819–829.
Rizopoulos, D., Hatfield, L.A., Carlin, B.P., Takkenberg, J.J., 2014. Combining dynamic predictions from joint models for longitudinal and time-to-event data using Bayesian model averaging. J. Amer. Statist. Assoc. 109 (508), 1385–1397.
Ruppert, D., 2002. Selecting the number of knots for penalized splines. J. Comput. Graph. Statist. 11 (4), 735–757.
Ruppert, D., Wand, M.P., Carroll, R.J., 2003. Semiparametric Regression, No. 12. Cambridge University Press.
Schmand, B., Huizenga, H.M., Gool, W.A.v., 2010. Meta-analysis of CSF and MRI biomarkers for detecting preclinical Alzheimer's disease. Psychol. Med. 40 (1), 135–145.
Tsiatis, A.A., Davidian, M., 2004. Joint modeling of longitudinal and time-to-event data: an overview. Statist. Sinica 14 (3), 809–834.
Wang, J., Luo, S., Li, L., 2017. Dynamic prediction for multiple repeated measures and event time data: An application to Parkinsons disease. Ann. Appl. Stat. 11 (3), 1787.
Weiner, M.W., Veitch, D.P., Aisen, P.S., Beckett, L.A., Cairns, N.J., Green, R.C., Harvey, D., Jack, C.R., Jagust, W., Liu, E., et al., 2013. The alzheimer's disease neuroimaging initiative: a review of papers published since its inception. Alzheimer's Dement. 9 (5), e111–e194.