Dynamic predictions in Bayesian functional joint models for longitudinal and time-to-event data: An application to Alzheimer's disease

Kan Li and Sheng Luo

Abstract

In the study of Alzheimer's disease, researchers often collect repeated measurements of clinical variables, event history, and functional data. If the health measurements deteriorate rapidly, patients may reach a level of cognitive impairment and are diagnosed as having dementia. An accurate prediction of the time to dementia based on the information collected is helpful for physicians to monitor patients' disease progression and to make early informed medical decisions. In this article, we first propose a functional joint model to account for functional predictors in both longitudinal and survival submodels in the joint modeling framework. We then develop a Bayesian approach for parameter estimation and a dynamic prediction framework for predicting the subjects' future outcome trajectories and risk of dementia, based on their scalar and functional measurements. The proposed Bayesian functional joint model provides a flexible framework to incorporate many features both in joint modeling of longitudinal and survival data and in functional data analysis. Our proposed model is evaluated by a simulation study and is applied to the motivating Alzheimer's Disease Neuroimaging Initiative study.

1 Introduction

Joint modeling of longitudinal and survival data is a popular framework to appropriately analyze datasets with repeated measurements and time-to-event outcomes.[1,2] The principle is to define two submodels (a mixed effects submodel for the longitudinal outcome and a Cox submodel for the survival outcome) and link them using a common latent structure. This modeling approach analyzes the two types of outcomes simultaneously and is able to reduce the bias of parameter estimates and improve the efficiency of statistical inference. A novel use of joint models, which has gained increasing interest in recent years, is to obtain dynamic personalized prediction of future longitudinal outcome trajectories and risks of survival events at any time, given the subject-specific outcome profiles up to the time of prediction. A key feature of the dynamic prediction frameworks is that the predictive measures can be dynamically updated as additional longitudinal measurements become available for the target subjects, providing instantaneous risk assessment, e.g. Rizopoulos[3] and Taylor et al.[4] The practical impact of these dynamic prediction tools can be dramatic for the neurodegenerative diseases, because they provide unique insight and valuable guidance for clinical decision making on treatment selection, patient prognosis, and counseling to facilitate a targeted treatment. However, current state-of-the-art joint models do not incorporate functional data, which are increasingly collected in public health and medical studies, to better understand many complex diseases.

Department of Biostatistics, The University of Texas Health Science Center at Houston, Houston, USA

Corresponding author:
Sheng Luo, Department of Biostatistics, The University of Texas Health Science Center at Houston, 1200 Pressler Street, Houston, TX 77030, USA.
Email: sheng.t.luo@uth.tmc.edu

Functional data consist of a sample of functions that provide information about curves, surfaces, or anything else varying over a continuum. These functions are usually defined on a one-dimensional time domain, such as growth curve data, heart rate monitor data, and electroencephalogram (EEG) data. A growing volume of functional data is also collected on higher dimensional domains such as magnetic resonance imaging (MRI), positron emission tomography (PET), and functional magnetic resonance imaging (fMRI). Functional data are closely related to multivariate data because functions are highly multivariate objects.[5] The main distinction between functional data and multivariate data are the natural ordering (in time or space) within a function. The high dimensionality and the complex structure in functional data pose challenges in both statistical theory and computation.[6,7]

Functional regression, especially functional predictor regression, is an active area of functional data analysis in the past decade. Although there is a rich literature in functional predictor regression to model the relationship between a scalar outcome and functional predictors,[8–13] most of the work focuses on cross-sectional data. Goldsmith et al.[14] proposed a penalized functional regression model to handle longitudinal measurements in both the response variable and functional predictors by incorporating scalar random effects. Gertheiss et al.[15] improved the longitudinal model to allow for different effects of subject-specific curves. More recently, Gellar et al.[16] extended the Cox proportional hazards model to incorporate functional predictors and estimated the parameters via penalized partial likelihood approach. Lee et al.[17] developed a Bayesian functional Cox regression model with both functional and scalar covariates, but used different regularization approaches. However, these works focus on the statistical inference instead of prediction. To the best of our knowledge, there are no studies of dynamic prediction based on functional regression modeling that simultaneously analyze the longitudinal measurements and time-to-event data. By using functional data as predictors in the dynamic prediction framework using a joint model, the rich information in functional data may increase the model's power of predicting disease progression in clinical practice.

In this article, we propose a novel joint model that incorporates the growing volume of functional data in the longitudinal-survival setting. Specifically, we develop a functional joint model (FJM), where outcomes consist of a longitudinal measure and a time-to-event variable, and the exposure variables include both scalar predictors and functional predictors. The key idea of characterizing FJM is to treat each functional predictor as a single structured object rather than a collection of data points,[7] in contrast to the scalar predictor. Thus, the FJM needs to handle the complex structure within each function and the association between different data types. In addition, the coefficient for a functional predictor in the FJM is also a function, which increases the difficulty for estimation and inference. We estimate the coefficient functions for the functional predictors using penalized spline approach. We develop a Bayesian approach for statistical inference of our FJM and a dynamic prediction framework for the predictions of target patients' future outcome trajectories and risks of event. These important predictive measures can provide valuable information to discover and validate prognostic biomarkers, which may advance the design of future clinical trials.

The rest 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 joint longitudinal-survival model with functional predictors, Bayesian inference procedure, and dynamic prediction framework. In Section 4, we apply the proposed method to the ADNI study. In Section 5, we conduct a simulation study to examine the performance of the proposed Bayesian FJM. Concluding remarks and discussion are presented in Section 6.

2 A motivating clinical study

The methodology development is motivated by the ADNI study. The primary goal of the study is to test whether serial MRI, positron emission tomography (PET), cerebrospinal fluid (CSF) markers, and neuropsychological assessments can be combined to measure the progression of Alzheimer's disease (AD). The phase 1 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.

Because MCI is commonly considered as a transitional stage between normal cognition and Alzheimer's disease, numerous recent studies assess various clinical markers and neuroimaging techniques to predict AD

diagnosis among MCI patients.[18] 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 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 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.

Moreover, the longitudinal Alzheimer Disease Assessment Scale-Cognitive (ADAS-Cog) score was reported to be the strongest predictor of time from MCI-to-AD.[19] It is an important clinical measure for cognitive functions and it manifests disease status. It is usually reported as a composite score of the 11 items (ADAS-Cog 11), with a total score of 70 and higher score indicating poor cognitive function. Figure 1 displays the lowess smoothing curve[20] of ADAS-Cog 11 scores over time for the MCI patients, with follow-up time less than 3 years (203 patients), 3–6 years (82 patients), and more than 3 years (70 patients), in addition to 95% pointwise confidence intervals. Figure 1 suggests that the ADAS-Cog 11 scores of patients in all three groups increase with time, which indicates the deteriorating of cognitive functions. The patients with shorter follow-up time tend to have higher ADAS-Cog 11 scores, indicating that patients with more severe cognitive impairment were more likely to progress to AD or censoring. This phenomenon manifests strong correlation between the longitudinal ADAS-Cog 11 values and the time to AD diagnosis, and is often referred to as "dependent censoring" or "informative censoring" in the literature of joint modeling.[21] However, many studies[22,23] designed to explore the association between longitudinal measures and disease progression of MCI patients analyzed the two processes separately. Their approaches failed to account for such informative censoring and resulted into biased inference.

Furthermore, the degree of atrophy within the medial temporal lobe structures, especially within the hippocampus, was reported to be associated with the AD progression.[24] Most of the current analysis was based on volumes of brain regions.[24–26] For example, AD patients and MCI patients have been shown to have 27% and 11% smaller hippocampal volumes, respectively, as compared with normal age-matched elderly.[24] However, some research[27,28] demonstrated that the surface-based morphology analysis offers more advantages because this method studies patterns of subregion atrophy and produce detailed pointwise correlation between atrophy and cognitive functions. In these surface-based analyses, 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 based on it. In this article, we propose a Bayesian personalized prediction model based on the FJM of longitudinal ADAS-Cog 11 score and the time to AD diagnosis, accounting for the clinical covariates and MRI measures. We include as a functional predictor the hippocampal radial distance (HRD) of
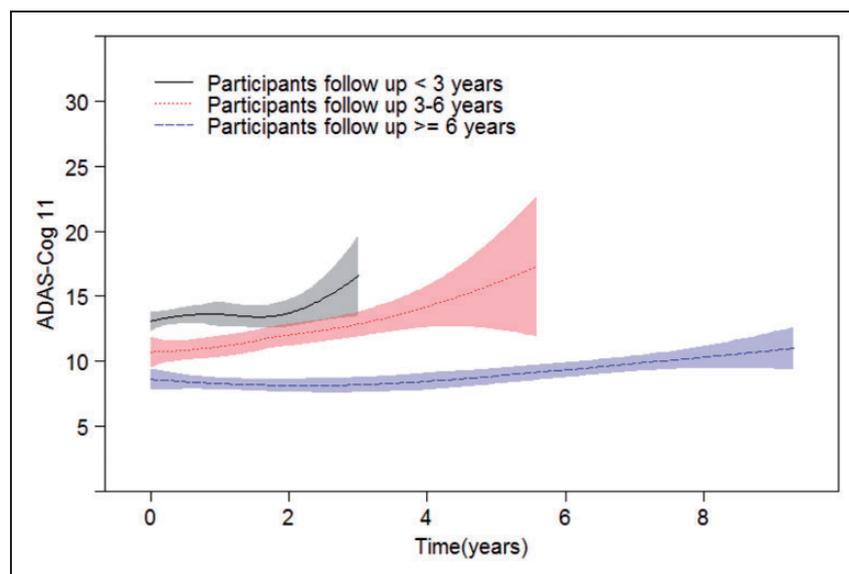

Figure 1. Smoothing curves of variable ADAS-Cog 11 over time for MCI patients with follow-up time less than 3 years (203 patients, solid line), 3–6 years (82 patients, dotted line), and more than 6 years (70 patients, dashed line). The shaded regions are 95% pointwise confidence intervals. ADAS-Cog: Alzheimer Disease Assessment Scale-Cognitive; MCI: mild cognitive impairment.

vertices on bilateral hippocampal surfaces. The HRD measures the distance from the medial core to each surface vertex and represents the hippocampal thickness. The HRD is calculated based on the surface model and the image processing procedure is detailed in the Web Supplement.

3 Methods

3.1 FJM framework

For each subject i ($i = 1, \ldots, I$) at visit j ($j = 1, \ldots, J_i$), we observe data $\{y_{ij}, \boldsymbol{x}_{ij}, g_i^{(x)}(s)\}$, where $y_{ij} = y_i(t_{ij})$ is a scalar outcome recorded at time t_{ij} from the study onset. Vector $\boldsymbol{x}_{ij}$ is a p-dimensional covariate vector. Function $g_i^{(x)}(s)$ is a time-invariant functional predictor defined over a 1D domain $s \in [0, S_{max}] = \boldsymbol{S}$. The domain of the functional predictor $\boldsymbol{S}$ is not the same as the time domain t over which the survival event is followed. For the ease of illustration, we only incorporate a single time-invariant functional predictor, such as the baseline MRI measure, in both the longitudinal and survival submodels, while the FJM can readily accommodate multiple functional predictors. We use superscript (x) and (w) to denote the functional predictors in the longitudinal and survival submodels, respectively. The longitudinal submodel is

$$y_i(t_{ij}) = m_i(t_{ij}) + \varepsilon_{ij}, \quad m_i(t_{ij}) = \beta_0 + \boldsymbol{x}_{ij}^\top \boldsymbol{\beta} + \int_{\boldsymbol{S}} g_i^{(x)}(s) B^{(x)}(s) \mathrm{d}s + \boldsymbol{z}_{ij}^\top \boldsymbol{u}_i \tag{1}$$

where $m_i(t_{ij})$ is the unobserved true value of the longitudinal outcome at time t_{ij}, β_0 is the intercept, and $\boldsymbol{\beta}$ is the regression coefficient vector. Coefficient function $B^{(x)}(s)$ (defined on the same domain as $g_i^{(x)}(s)$) determines a pointwise association between $g_i^{(x)}(s)$ and $y_i(t_{ij})$. Vector $\boldsymbol{z}_{ij}$ is a q-dimensional covariates corresponding to random effects $\boldsymbol{u}_i$, which is assumed to have $\boldsymbol{u}_i \sim N(0, \Sigma_u)$ to account for the within-subject correlation. The measurement error $\varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2)$ is independent from $\boldsymbol{u}_i$. To allow flexibility and smoothness in modeling the effects of some covariates, smooth functions using splines can also be included in model (1).

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

$$h_i(t) = h_0(t) \exp\left\{ \boldsymbol{w}_i^\top \gamma + \int_{\boldsymbol{S}} g_i^{(w)}(s) B^{(w)}(s) \mathrm{d}s + \alpha m_i(t) \right\} \tag{2}$$

where $h_0(t)$ is the baseline hazard function, and $\boldsymbol{w}_i$ is a vector of time-independent covariates with regression coefficient vector γ. Functional predictor $g_i^{(w)}(s)$ may be the same or different from its counterpart $g_i^{(x)}(s)$ in model (1). Functional log hazard ratio $B^{(w)}(s)$ measures the overall contribution of $g_i^{(w)}(s)$ towards the event hazard. The association parameter α quantifies the strength of correlation between the unobserved true longitudinal function $m_i(t)$ and the event hazard at the same time point t. Models (1) and (2) consist of the FJM framework.

To build the functional regression model, we adopt a penalized approach to incorporate functional components into functional predictor regression model.[14,16] We first express the time-invariant functional predictor $g_i^{(x)}(s)$ in model (1) using the Karhunen–Loève decomposition. Let $\mu^{(x)}(s)$ be the mean of $g_i^{(x)}(s)$ and $\Sigma^{(x)}(s, s') = cov\{g_i^{(x)}(s), g_i^{(x)}(s')\}$ be the covariance function between two locations (s and s') of the functional predictor. The spectral decomposition of the covariance function is given by $\Sigma^{(x)}(s, s') = \sum_{l=1}^\infty \lambda_l^{(x)} \phi_l^{(x)}(s) \phi_l^{(x)}(s')$, where $\lambda_1^{(x)} \geq \lambda_2^{(x)} \geq \cdots \geq 0$ are non-increasing eigenvalues and $\phi_l^{(x)}(s)$'s are the corresponding orthonormal eigenfunctions. The Karhunen–Loève expansion of $g_i^{(x)}(s)$ is $g_i^{(x)}(s) = \mu^{(x)}(s) + \sum_{l=1}^\infty \xi_{il}^{(x)} \phi_l^{(x)}(s)$, where the functional principal component (FPC) scores $\xi_{il}^{(x)} = \int_{\boldsymbol{S}} \{g_i^{(x)}(s) - \mu^{(x)}(s)\} \phi_l^{(x)}(s) \mathrm{d}s$ are uncorrelated random variables with mean zero and variance $\lambda_l^{(x)}$. In practice, we adopt a truncated approximation for $g_i^{(x)}(s)$ given by $g_i^{(x)}(s) \approx \mu^{(x)}(s) + \sum_{l=1}^{K_x} \xi_{il}^{(x)} \phi_l^{(x)}(s)$.

In the second step, we express coefficient function $B^{(x)}(s)$ in model (1) in term of a cubic B-spline basis functions $\boldsymbol{\psi}^{(x)}(s) = [\psi_1^{(x)}(s), \ldots, \psi_{K_B}^{(x)}(s)]^\top$, so that $B^{(x)}(s) = \sum_{l=1}^{K_B} B_l^{(x)} \psi_l^{(x)}(s)$. The basis function $\boldsymbol{\psi}(s)^{(x)}$ is evaluated over the domain $\boldsymbol{S}$ on which function $g_i^{(x)}(s)$ is observed. The coefficients $B_l^{(x)}$ are unknown parameters that need to be estimated. We denote the vector of FPC scores as $\boldsymbol{\xi}_i^{(x)} = [\xi_{i1}^{(x)}, \ldots, \xi_{iK_x}^{(x)}]^\top$, the vector of eigenfunctions as

$\phi^{(x)}(s) = [\phi_1^{(x)}(s), \ldots, \phi_{K_x}^{(x)}(s)]^\top$, and the vector of coefficients as $\boldsymbol{B}^{(x)} = [B_1^{(x)}, \ldots, B_{K_B}^{(x)}]^\top$. Then we have

$$\int_S g_i^{(x)}(s)B^{(x)}(s)\mathrm{d}s \approx \int_S \mu^{(x)}(s)B^{(x)}(s)\mathrm{d}s + \int_S \left(\xi_i^{(x)}\right)^\top \phi^{(x)}(s)\psi^{(x)}(s)^\top \boldsymbol{B}^{(x)}\mathrm{d}s$$

$$= \int_S \mu^{(x)}(s)B^{(x)}(s)\mathrm{d}s + \left(\xi_i^{(x)}\right)^\top \int_S \phi^{(x)}(s)\psi^{(x)}(s)^\top \mathrm{d}s \boldsymbol{B}^{(x)}$$

$$= \int_S \mu^{(x)}(s)B^{(x)}(s)\mathrm{d}s + \left(\xi_i^{(x)}\right)^\top \boldsymbol{J}_{\phi,\psi}^{(x)} \boldsymbol{B}^{(x)}$$

Note that $\boldsymbol{J}_{\phi,\psi}^{(x)}$ is a $K_x \times K_B$ matrix with the (k,l)th entry equal to $\int_S \phi_k^{(x)}(s)\psi_l^{(x)}(s)^\top \mathrm{d}s$. The integrals are computed by numeric integration and $\boldsymbol{J}_{\phi,\psi}^{(x)}$ is a fixed term in the model. The number of components K_x can be determined using the proportion of explained variance (PEV). Specifically, K_x may be chosen as the minimum number of FPCs such that $\sum_{l=1}^{K_x} \hat{\lambda}_l^{(x)} / \sum_{l=1}^{\infty} \hat{\lambda}_l^{(x)} \geq L$, where L is a pre-specified PEV, e.g. $L = 80\%, 90\%$, or 95%. However, the selected number K_x can explain the majority of the variability in the functional predictor, but it may not adequately represent the coefficient function. Thus, in this article, we refer to Ruppert[29] and choose K_x and K_B sufficient large (e.g. 20) to capture the complexity in both functional predictors and coefficient functions, with the identifiability constrain $K_x \geq K_B$.

Similarly, the functional predictor $g^{(w)}(s)$ in model (2) can be expressed as $g_i^{(w)}(s) \approx \mu^{(w)}(s) + \sum_{l=1}^{K_w} \xi_{il}^{(w)}\phi_l^{(w)}(s)$ and the coefficient function $B^{(w)}(s) = \sum_{l=1}^{K_B} B_l^{(w)}\psi_l^{(w)}(s)$, where $\psi_l^{(w)}(s)$'s are B-spline basis functions defined on the same domain of $g^{(w)}(s)$. Thus, $\int_S g^{(w)}(s)B^{(w)}(s)\mathrm{d}s \approx \int_S \mu^{(w)}(s)B^{(w)}(s)\mathrm{d}s + (\xi_i^{(w)})^\top \boldsymbol{J}_{\phi,\psi}^{(w)} \boldsymbol{B}^{(w)}$, with $\mu^{(w)}(s)$, $\xi_{il}^{(w)}$, $\phi_l^{(w)}(s)$, $\boldsymbol{J}_{\phi,\psi}^{(w)}$, and $\boldsymbol{B}^{(w)}$ have the same meanings as $\mu^{(x)}(s)$, $\xi_{il}^{(x)}$, $\phi_l^{(x)}(s)$, $\boldsymbol{J}_{\phi,\psi}^{(x)}$, and $\boldsymbol{B}^{(x)}$, respectively. Thus, the FJM is rewritten as

$$y_i(t) = m_i(t) + \varepsilon_{ij}, \quad \text{where } m_i(t) = \beta_0' + \boldsymbol{x}_{ij}^\top \boldsymbol{\beta} + (\xi_i^{(x)})^\top \boldsymbol{J}_{\phi,\psi}^{(x)} \boldsymbol{B}^{(x)} + \boldsymbol{z}_{ij}^\top \boldsymbol{u}_i \tag{3}$$

$$\text{and} \quad h_i(t) = h_0^*(t) \exp\left\{\boldsymbol{w}_i^\top \gamma + \left(\xi_i^{(w)}\right)^\top \boldsymbol{J}_{\phi,\psi}^{(w)} \boldsymbol{B}^{(w)} + \alpha m_i(t)\right\} \tag{4}$$

where $\beta_0' = \beta_0 + \int_S \mu^{(x)}(s)B^{(x)}(s)\mathrm{d}s$ and $h_0^*(t) = h_0(t)\exp\{\int_S \mu^{(w)}(s)B^{(w)}(s)\mathrm{d}s\}$. For notational ease and without ambiguity, we replace the approximation sign ($\approx$) by the equal sign. Note that models (3) and (4) are similar to a linear mixed submodel for the longitudinal scalar response variable and a Cox submodel for the survival outcome, respectively, in a standard joint model framework.[2] And FPC scores $\xi_i^{(x)}$ and $\xi_i^{(w)}$ can be treated as scalar covariates. Similar to mixed models, our FJM can readily handle unbalanced data in the longitudinal measurement of $y_i(t)$.

Let $\theta = [\beta_0', \boldsymbol{\beta}^\top, (\boldsymbol{B}^{(x)})^\top, \text{vech}(\Sigma_u)^\top, \gamma^\top, (\boldsymbol{B}^{(w)})^\top, \alpha, \sigma_\varepsilon^2, \theta_{h_0^*}^\top]^\top$ be the parameter vector, where $\text{vech}(\Sigma_u)$ is the vector being formed by vectorizing the lower triangular part of covariance matrix Σ_u, vector $\theta_{h_0^*}$ denotes the parameters in the baseline hazard function $h_0^*(\cdot)$. The conditional likelihood from the longitudinal data $\boldsymbol{y}_i = [y_{i1}, \ldots, y_{iJ_i}]^\top$ is

$$p(\boldsymbol{y}_i|\theta, \boldsymbol{u}_i) = (2\pi\sigma_\varepsilon^2)^{-J_i/2} \exp\left\{-\frac{1}{2\sigma_\varepsilon^2}\sum_{j=1}^{J_i}\left[y_{ij} - \left(\beta_0' + \boldsymbol{x}_{ij}^\top \boldsymbol{\beta} + \left(\xi_i^{(x)}\right)^\top \boldsymbol{J}_{\phi,\psi}^{(x)} \boldsymbol{B}^{(x)} + \boldsymbol{z}_{ij}^\top \boldsymbol{u}_i\right)\right]^2\right\}$$

and the density function of the random effects $\boldsymbol{u}_i$ is $p(\boldsymbol{u}_i|\theta) = (2\pi)^{-q/2}|\Sigma_u|^{-1/2}\exp(-\frac{1}{2}\boldsymbol{u}_i^\top \Sigma_u^{-1}\boldsymbol{u}_i)$, where q is the dimension of the covariance matrix Σ_u. The conditional likelihood from the survival data are

$$p(T_i, \delta_i, |\theta, \boldsymbol{u}_i) = h_i(T_i|\theta, \boldsymbol{u}_i)^{\delta_i} S_i(T_i|\theta, \boldsymbol{u}_i) = h_i(T_i|\theta, \boldsymbol{u}_i)^{\delta_i}\exp\left[-\int_0^{T_i} h_i(t|\theta, \boldsymbol{u}_i)\mathrm{d}t\right]$$

where $h_i(T_i|\theta, \boldsymbol{u}_i) = h_0^*(T_i)\exp\{\boldsymbol{w}_i^\top \gamma + (\xi_i^{(w)})^\top \boldsymbol{J}_{\phi,\psi}^{(w)} \boldsymbol{B}^{(w)} + \alpha m_i(T_i)\}$, 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 effect vector $\boldsymbol{u}_i$, all components in $\boldsymbol{y}_i$ and T_i are independent), the joint likelihood function is

$$L(\theta) = \prod_{i=1}^{I} p(\boldsymbol{y}_i, T_i, \delta_i|\theta) = \prod_{i=1}^{I} \int p(\boldsymbol{y}_i|\theta, \boldsymbol{u}_i)\, p(T_i, \delta_i, |\theta, \boldsymbol{u}_i)\, p(\boldsymbol{u}_i|\theta)\mathrm{d}\boldsymbol{u}_i \tag{5}$$

To prevent overfitting, we adopt penalization technique to estimate spline coefficients $\boldsymbol{B}^{(x)}$ and $\boldsymbol{B}^{(w)}$ and to introduce smoothness in the resulting coefficient function. The penalized log-likelihood is formed by subtracting the penalty terms from the log-likelihood,

$$l(\boldsymbol{\theta}) = \log L(\boldsymbol{\theta}) - \frac{1}{\sigma_{B^{(x)}}^2}(\boldsymbol{B}^{(x)})^\top \boldsymbol{B}^{(x)} - \frac{1}{\sigma_{B^{(w)}}^2}(\boldsymbol{B}^{(w)})^\top \boldsymbol{B}^{(w)} \tag{6}$$

where $\sigma_{B^{(x)}}^2$ and $\sigma_{B^{(w)}}^2$ are smoothing parameters. Instead of fitting the model by direct maximization of the penalized likelihood, we adopt a Bayesian penalization approach detailed in Section 3.2.

3.2 Bayesian inference

In practice, the functional predictors such as $g_i^{(x)}(s)$ are measured over finite grids in domain $\boldsymbol{S}$ and often with measurement error, i.e. the observed functional predictor $g_i^{(x')}(s) = g_i^{(x)}(s) + \epsilon_i(s)$, where measurement error $\epsilon_i(s) \sim N(0, \sigma_\epsilon^2)$ is a white noise process. The mean function $\mu^{(x)}(s)$ is estimated by $\hat{\mu}^{(x)}(s) = \sum_{i=1}^I g_i^{(x')}(s)/I$, and the empirical covariance function is estimated by $\hat{\Sigma}^{(x)}(s,s') = \sum_{i=1}^I \{g_i^{(x')}(s) - \hat{\mu}^{(x)}(s)\}\{g_i^{(x')}(s') - \hat{\mu}^{(x)}(s')\}/(I-1)$. We apply kernel smoothing to the off-diagonal elements of $\hat{\Sigma}_x(s,s')$ to remove the effects from measurement errors.[30,31] Then the estimated eigenvalues $\hat{\lambda}_l^{(x)}$ and the corresponding estimated eigenfunctions $\hat{\phi}_l^{(x)}(s)$, where $l = 1, \ldots, S_{max}$, are calculated based on the decomposition of the smoothed covariance function. Finally, the estimated FPC scores for each subject are calculated as $\hat{\xi}_{il}^{(x)} = \int_{\boldsymbol{S}} \{g_i^{(x')}(s) - \hat{\mu}^{(x)}(s)\}\hat{\phi}_l^{(x)}(s)\mathrm{d}s$, and the integral can be approximated by the Riemann sum. We choose the first K_x estimated eigenfunctions and FPC scores, and denote them as $\widehat{\phi}^{(x)}(s) = [\hat{\phi}_1^{(x)}(s), \ldots, \hat{\phi}_{K_x}^{(x)}(s)]^\top$ and $\widehat{\xi}_i^{(x)} = [\hat{\xi}_{i1}^{(x)}, \ldots, \hat{\xi}_{iK_x}^{(x)}]^\top$, respectively. In a similar fashion, functional predictor $g_i^{(w)}(s)$ can be obtained and the components $\widehat{\phi}^{(w)}(s)$ and $\widehat{\xi}_i^{(w)}$ can be defined.

For model fitting, we propose a Bayesian approach based on Markov Chain Monte Carlo (MCMC) posterior simulations, which provides a flexible way for statistical inference. In Bayesian framework, unknown parameters are considered as random variables with appropriate prior distributions. We use vague prior distributions on all elements in parameter vector $\boldsymbol{\theta}$. Specifically, the prior distributions of parameters $\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\alpha}$ are $N(0, 100)$, and Inverse-Gamma $(0.01, 0.01)$ for all variance parameters. We impose smoothness on coefficient function estimates through the prior specification on $B^{(x)}$ and $B^{(w)}$. We replace the penalty terms in the penalized likelihood (6) by their stochastic analogues, and assume a random walk prior distribution on the $B_l^{(x)}$ and $B_l^{(w)}$, for $l = 1, \ldots, K_B$.[32] Specifically, we use a first-order random prior distribution for $B_{l+1}^{(x)} \sim N(B_{l+1}^{(x)}, \sigma_{B^{(x)}}^2)$, for $l = 1, \ldots, K_B - 1$, where $B_1^{(x)}$ is treated as a fixed unknown parameter. Then the variance component $\sigma_{B^{(x)}}^2$ is assigned Inverse-Gamma $(0.01, 0.01)$ as prior distribution. A similar approach is applied to $B_l^{(w)}$. We have investigated other selections of prior distributions and hyper parameters and have obtained very similar results.

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,[33] which 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.[34] After fitting the model to the training dataset (the dataset used to build the model) using Bayesian approaches, we obtain D (e.g. $D = 2000$ 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 $\widehat{\boldsymbol{B}}^{(x)}$ (posterior mean), the estimated coefficient function is calculated by $\hat{B}^{(x)}(s) = (\widehat{\boldsymbol{B}}^{(x)})^\top \boldsymbol{\psi}^{(x)}(s)$. The Bayesian approach allows for the easy construction of posterior credible intervals for the coefficient function $B^{(x)}(s)$ as $[\hat{q}_{B,0.025}(s), \hat{q}_{B,0.975}(s)]$, where $\hat{q}_{B,p}(s)$ is the p-quartile of the MCMC samples $[B^{(x)}(s)]^{(d)} = [(\boldsymbol{B}^{(x)})^{(d)}]^\top \boldsymbol{\psi}^{(x)}(s), d = 1, \ldots, D$. Similarly, the estimated coefficient function $\hat{B}^{(w)}(s)$ and its credible interval can be obtained. To facilitate easy reading and implementation of the proposed FJM, a Stan code has been posted in the Web Supplement.

3.3 Dynamic prediction framework

We next illustrate the dynamic prediction framework based on the proposed Bayesian FJM. Given a new subject N's outcome histories $\boldsymbol{y}_N^{\{t\}} = \{y_N(t_{Nj}); 0 \leq t_{Nj} \leq t\}$ and covariates $X_N^{\{t\}} = \{\boldsymbol{x}_N(t_{Nj}), g_N^{(x)}(s), \boldsymbol{w}_N, g_N^{(w)}(s); 0 \leq t_{Nj} \leq t\}$ up to time t, and $\delta_N = 0$ (no event), we want to predict the personalized longitudinal outcome $y_N(t')$ at a future time point $t' > t$ (e.g. $t' = t + \Delta t$), and the conditional probability of event-free or survival at time t', denoted by $\pi_N(t'|t) = P(T_N^* \geq t' | T_N^* > t, \boldsymbol{y}_N^{\{t\}}, X_N^{\{t\}})$. The key step for prediction is to obtain samples for subject N's random effect vector $\boldsymbol{u}_N$ from its posterior distribution $p(\boldsymbol{u}_N | T_N^* > t, \boldsymbol{y}_N^{\{t\}}, \boldsymbol{\theta})$. Conditional on the dth posterior sample $\boldsymbol{\theta}^{(d)}$, $d = 1, \ldots, D$, we draw the dth sample of the random effect vector $\boldsymbol{u}_N$ from its posterior distribution

$$p\left(\boldsymbol{u}_N | T_N^* > t, \boldsymbol{y}_N^{\{t\}}, \boldsymbol{\theta}^{(d)}\right) = \frac{p\left(\boldsymbol{y}_N^{\{t\}}, T_N^* > t, \boldsymbol{u}_N | \boldsymbol{\theta}^{(d)}\right)}{p\left(\boldsymbol{y}_N^{\{t\}}, T_N^* > t | \boldsymbol{\theta}^{(d)}\right)} \propto p\left(\boldsymbol{y}_N^{\{t\}}, T_N^* > t, \boldsymbol{u}_N | \boldsymbol{\theta}^{(d)}\right)$$

$$= p\left(\boldsymbol{y}_N^{\{t\}} | \boldsymbol{\theta}^{(d)}, \boldsymbol{u}_N\right) p\left(T_N^* > t | \boldsymbol{\theta}^{(d)}, \boldsymbol{u}_N\right) p\left(\boldsymbol{u}_N | \boldsymbol{\theta}^{(d)}\right)$$

where $p(\boldsymbol{y}_N^{\{t\}} | \boldsymbol{\theta}^{(d)}, \boldsymbol{u}_N)$ is the conditional probability of longitudinal data from model (3), $p(T_N^* > t | \boldsymbol{\theta}^{(d)}, \boldsymbol{u}_N)$ is the survival probability from model (4), and $p(\boldsymbol{u}_N | \boldsymbol{\theta}^{(d)})$ is the probability of random effect. For each of $\boldsymbol{\theta}^{(d)}$, $d = 1, \ldots, D$, we use adaptive rejection Metropolis sampling (ARMS)[35] to draw one sample of random effect vector $\boldsymbol{u}_N$. This process is repeated for the D saved values of $\boldsymbol{\theta}$, so that D samples of random effect vector $\boldsymbol{u}_N$ are obtained. The predictions can be calculated by plugging in the samples of the parameter vector and random effect vector $\{\boldsymbol{\theta}^{(d)}, \boldsymbol{u}_N^{(d)}, d = 1, \ldots, D\}$ into the proposed models. For example, based on model (3), the expected values of the longitudinal outcome for subject N at time t' is

$$\hat{E}\left\{y_N(t') | T_N^* > t, \boldsymbol{y}_N^{\{t\}}, X_N^{\{t\}}\right\} = \int E\left\{y_N(t') | T_N^* > t, \boldsymbol{y}_N^{\{t\}}, X_N^{\{t\}}, \boldsymbol{u}_N\right\} p\left(\boldsymbol{u}_N | T_N^* > t, \boldsymbol{y}_N^{\{t\}}, X_N^{\{t\}}\right) d\boldsymbol{u}_N$$

$$\approx \frac{1}{D} \sum_{d=1}^{D} E\left\{y_N(t') | T_N^* > t, \boldsymbol{y}_N^{\{t\}}, X_N^{\{t\}}, \boldsymbol{u}_N^{(d)}\right\} = \frac{1}{D} \sum_{d=1}^{D} m_N^{(d)}(t')$$

where the integration with respect to $\boldsymbol{u}_N$ in the first equality is approximated using Monte Carlo method. The baseline functional predictor such as $g_N^{(x)}(s)$ can be converted to $g_N^{(x)}(s) \approx \hat{\mu}^{(x)}(s) + \sum_{l=1}^{K_x} \hat{\xi}_{Nl}^{(x)} \hat{\phi}_l^{(x)}(s)$, where $\hat{\xi}_{Nl}^{(x)} = \int_S \{g_N^{(x)}(s) - \hat{\mu}^{(x)}(s)\} \hat{\phi}_l^{(x)}(s) ds$. Similarly, based on model (4), the conditional probability of event-free at time t' is

$$\hat{\pi}_N(t'|t) = \int P(T_N^* \geq t' | T_N^* > t, \boldsymbol{y}_N^{\{t\}}, X_N^{\{t\}}, \boldsymbol{u}_N) p(\boldsymbol{u}_N | T_N^* > t, \boldsymbol{y}_N^{\{t\}}, X_N^{\{t\}}) d\boldsymbol{u}_N$$

$$\approx \frac{1}{D} \sum_{d=1}^{D} P\left(T_N^* \geq t' | T_N^* > t, \boldsymbol{y}_N^{\{t\}}, X_N^{\{t\}}, \boldsymbol{u}_N^{(d)}\right) = \frac{1}{D} \sum_{d=1}^{D} \frac{P\left(T_N^* \geq t' | T_N^* > t, \boldsymbol{y}_N^{\{t\}}, X_N^{\{t\}}, \boldsymbol{u}_N^{(d)}\right)}{P\left(T_N^* \geq t | T_N^* > t, \boldsymbol{y}_N^{\{t\}}, X_N^{\{t\}}, \boldsymbol{u}_N^{(d)}\right)}$$

$$= \frac{1}{D} \sum_{d=1}^{D} \exp\left[-\int_t^{t'} h_N^{(d)}\left(t^* | \boldsymbol{u}_N^{(d)}\right) dt^*\right]$$

Suppose that subject N has not experienced the event of interest by time t', then the outcome histories are updated to $y_N^{\{t'\}}$. We can dynamically update the posterior distribution to $p(\boldsymbol{u}_N | T_N^* > t', \boldsymbol{y}_N^{\{t'\}}, \boldsymbol{\theta})$, draw new samples, and obtain the updated predictions.

It is essential to assess the performance of the proposed predictive measures. Information criteria, such as deviance information criterion (DIC), can be useful to assess the overall predictive ability of the model. Here we mainly focus on the survival outcome, probability of event-free $\pi_N(t'|t)$, and on how well the model discriminates between patients who had the event from patients who did not. Such discrimination performance is measured by the time-dependent receiver operating characteristic (ROC) curve and the area under the ROC curve (AUC; higher value indicates better discrimination performance). Specifically, for any given cut point $c \in (0, 1)$, the time-dependent sensitivity and specificity are defined as sensitivity: $P(\pi_i(t'|t) > c | N_i(t, t') = 1, T_i^* > t)$ and specificity: $P(\pi_i(t'|t) \leq c | N_i(t, t') = 0, T_i^* > t)$, respectively, where $N_i(t, t') = I(t < T_i^* \leq t')$ indicating whether there is an observed event or not for subject i during the time interval $(t, t']$. In the absence of censoring, sensitivity and specificity can be simply estimated from the empirical distribution of the predicted risk. To accommodate

censoring time, Li et al.[36] proposed a kernel weighting method to estimate the time-dependent ROC curve nonparametrically and a closed-form formula to calculate ROC. Consistent estimation of the time-dependent sensitivity and specificity can be achieved by weighting the patients by their conditional probabilities of disease onset prior to the time horizon, given the data. This method produces unbiased and efficient estimators and is less sensitive to the selection of kernel bandwidth as compared with Heagerty et al.[37]

Moreover, we also assess our model's performance in calibration, i.e. the agreement between the predicted and true risks. We use the dynamic expected Brier score (BS) for joint models, which is an extension from the BS developed in the survival model.[38,39] The dynamic expected BS is defined as $E[(D(t'|t) - \pi(t'|t))^2]$, where the observed failure status $D(t'|t)$ equals to 1 if the subject experiences the terminal event within the time interval $(t, t']$ and 0 if the subject is event free until t'. An estimator of BS is

$$\widehat{BS}(''t'|t'') = \frac{1}{N_t} \sum_{i=1}^{N_t} \hat{G}_i(''t'|t'') \left(D_i(''t'|t'') - \pi_i(t'|t) \right)^2$$

where N_t is the number of subjects at risk at time t, and the weight $\hat{G}_i(''t'|t'') = \frac{I(t_i > t')}{\hat{S}_0(t')/\hat{S}_0(t)} + \frac{I(t < t_i \leq t')\delta_i}{\hat{S}_0(t_i)/\hat{S}_0(t)}$ is to account for censoring with $\hat{S}_0$ denoting the Kaplan–Meier estimate.[39] In general, $BS = 0$ indicates perfect prediction and $BS = 0.25$ means no better than a random guess.

3.4 Implementation using software

An advantage of the proposed Bayesian FJM is that its implementation and extension can be done via available softwares and the sample `Stan` code. The first step is to conduct functional principal component analysis (FPCA) for functional predictors and estimate the FPC eigenfunctions and scores, by using either `fpca.sc` function in the `refund` packages[40] in R or `fpca.mle` and `fpca.score` functions in the FPCA package[41] in R. In the second step, the cubic B-spline basis functions can be evaluated over the discretized domain S using the `bs` function in the package `spline` in R. In the third step, the FPC scores obtained from FPCA are used as scalar covariates in the joint model and the parameters can be estimated using the sample `Stan` code. For prediction, the samples of the random effect vector $\boldsymbol{u}_N$ can be drawn from its posterior distribution using the `arms` function in the `HI` package in R. The future survival probability can then be calculated based on the samples. Finally, assessment of the predictive performance based on time-dependent ROC can be achieved using the `tdROC` package in R.

4 Application to the ADNI study

We apply the proposed Bayesian FJM to the motivating ADNI-1 study. 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 allele ($APOE - \varepsilon4$, 56%), given their potential effects on AD progression.[25,26,42] To utilize the brain imaging information, we include baseline hippocampal volume (*bHV*) as a scalar covariate and the baseline hippocampal surface based on *HRD* as a functional predictor. We follow the procedure in Section 2 to convert the 3D *HRD* to a 1D domain denoted by S. In an exploratory analysis of the ADNI data, we plot the lowess curve of the longitudinal outcome ADAS-Cog 11 against each scalar covariate and there is no strong nonlinearity in these curves. Hence, we include linear terms of these covariates, and avoid using splines to have easy interpretation and less computational burden.

The first model we consider is the regular joint model (refer to as model *JM*), which incorporates variable *bHV* in both longitudinal and survival submodels. Additionally, we consider three *FJMs*, i.e. model *FJM*1 that includes *HRD* only in the longitudinal submodel, model *FJM*2 that includes *HRD* only in the survival submodel, and model *FJM*3 that includes *HRD* in both submodels as

$$ADAS - Cog_i(t_{ij}) = m_i(t_{ij}) + \varepsilon_{ij}$$

$$m_i(t_{ij}) = \beta_0 + \beta_1 t_{ij} + \beta_2 bAge_i + \beta_3 bHV_i + \int_S HRD_i(s)B^{(x)}(s)\mathrm{d}s + u_{i1}$$

$$h_i(t) = h_0(t)\exp\{\gamma_1 gender_i + \gamma_2 bAge_i + \gamma_3 Edu_i + \gamma_4 APOE - \varepsilon4$$

$$+ \gamma_5 bHV_i + \int_S HRD_i(s)B^{(w)}(s)\mathrm{d}s + \alpha m_i(t)\}$$

Table 1. AUCs and BSs by four candidate models in the ADNI study.

Δt	t	JM		FJM1		FJM2		FJM3	
		AUC	BS	AUC	BS	AUC	BS	AUC	BS
0.5	1	0.818	0.086	0.818	0.086	0.815	0.084	0.814	0.084
	1.5	0.728	0.118	0.729	0.118	0.755	0.112	0.778	0.109
	2	0.825	0.103	0.831	0.102	0.877	0.087	0.875	0.086
1	1	0.784	0.150	0.785	0.150	0.793	0.145	0.808	0.142
	1.5	0.757	0.171	0.762	0.170	0.812	0.150	0.823	0.146
	2	0.741	0.142	0.746	0.139	0.789	0.129	0.792	0.128

AUC: Area under the ROC curve; BS: Brier score; ADNI: Alzheimer's Disease Neuroimaging Initiative; JM: joint model; FJM: functional joint model.

We perform FPCA on *HRD* and select the first 20 FPCs, along with their FPC scores. We expand the coefficient functions $B^{(x)}(s)$ and $B^{(w)}(s)$ using the same cubic B-spline basis functions $\psi(s)^{(x)} = \psi(s)^{(w)}$ with $K_B = 20$. Baseline hazard function $h_0(t)$ is approximated by a piecewise constant function. Specifically, the observed survival time is divided into $M = 7$ intervals by every $1/M$th quantiles. We have also explored other selections of M and obtained very similar results.

We compare the four candidate models by 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. Specifically, the total sample of patients is randomly splitted into 10 subgroups of about equal size. The analysis is repeated 10 times with one subset being set aside as the validation dataset and the remaining nine subsets being used as the training dataset. 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 Δt as 0.5 and 1 years for analysis. Then the time-dependent AUCs and BSs are calculated based on the predicted probabilities of all patients.

Table 1 displays the time-dependent AUCs and dynamic expected BSs from the four candidate models. Model *FJM*2 and *FJM*3 have notably larger AUCs and smaller BSs than models *JM* and *FJM*1 for most of combinations of t and Δt. This suggests that including functional predictor *HRD* in the survival submodel, in addition to scalar predictor hippocampal volume, improves the capability of the joint model in predicting risk of AD diagnosis. We also notice that *FJM*3 has a smaller DIC than *FJM*2 (12063 compares to 12097), which indicates that *HRD* is an important functional predictor for the cognitive functions manifested by variable ADAS-Cog 11 among the MIC patients. These results suggest that including *HRD* in the longitudinal submodel may further improve the overall predictive ability of the FJM model. Hence, we select *FJM*3 as the final model because it has a competitive good discrimination capability and a smaller DIC value.

Parameter estimates from model *FJM*3 using whole dataset are presented in Table 2, while the estimated vector of coefficients $\widehat{\boldsymbol{B}}^{(x)}$ and $\widehat{\boldsymbol{B}}^{(w)}$ (for a vector of 20 FPC scores $\boldsymbol{\xi}_i^{(x)}$ as in models (3) and (4)) are presented in Web Tables 1 and 2. In the longitudinal submodel, the ADAS-Cog 11 score increases (deteriorates) as time progresses, i.e. an average increase of 0.428 unit (95% CI: [0.338–0.521]) per year for MCI patients. Larger hippocampal volume at baseline (*bHV*) is associated with lower (better) ADAS-Cog 11 scores. In the survival submodel, the presence of *APOE*-ε4 allele(s) increases the hazard of AD diagnosis by 70% (exp(0.533) − 1, 95% CI: [26%–107%]), which is consistent with the literature.[43] Older age at baseline is associated with lower risk of AD diagnosis. The hippocampal volume (*bHV*) is no longer significant after including the functional predictor HRD in the survival submodel. 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 14.3% (exp(0.134) − 1, 95% CI: [8%–19%]).

The coefficient function of $B^{(x)}(s)$ for *HRD* in the longitudinal submodel is estimated via $\hat{B}^{(x)}(s) = \sum_{l=1}^{20} \hat{B}_l^{(x)} \psi_l^{(x)}(s)$, and $B^{(w)}(s)$ in the survival submodel can be estimated in the same way. For visualization purpose, we map $\hat{B}^{(x)}(s)$ and $\hat{B}^{(w)}(s)$ back to the corresponding vertex on the hippocampal surfaces (Figure 2). Due to the difficulty of displaying a 3D object on paper, Figure 2 only displays two views (from top and bottom) of left and right hippocampal surfaces. Panel (a) displays a schematic representation of the hippocampal subfields, defined by Apostolova et al.,[44] on the hippocampal surface template. Panel (b) displays the coefficient function

Table 2. ADNI data analysis results from model *FJM*3.

	Parameters	Mean	SE	2.5%	97.5%
For longitudinal outcome					
ADAS-Cog 11	Time (years)	0.428	0.045	0.338	0.521
	bAge	−0.364	0.260	−0.885	0.156
	bHV (mm^3)	−1.617	0.295	−2.201	−1.051
For survival process					
MCI to AD	Female	−0.088	0.173	−0.397	0.270
	bAge	−0.283	0.042	−0.423	−0.109
	Edu (years)	0.028	0.016	−0.002	0.062
	APOE-ε	0.533	0.125	0.239	0.728
	bHV (mm^3)	0.056	0.114	−0.185	0.276
	α	0.134	0.022	0.079	0.177

ADNI: Alzheimer's Disease Neuroimaging Initiative; FJM: functional joint model; SE: standard error; ADAS-Cog: Alzheimer Disease Assessment Scale-Cognitive; *bAge*: baseline age; *bHV*: baseline hippocampal volume; MCI: mild cognitive impairment; AD: Alzheimer's disease; Edu: education.
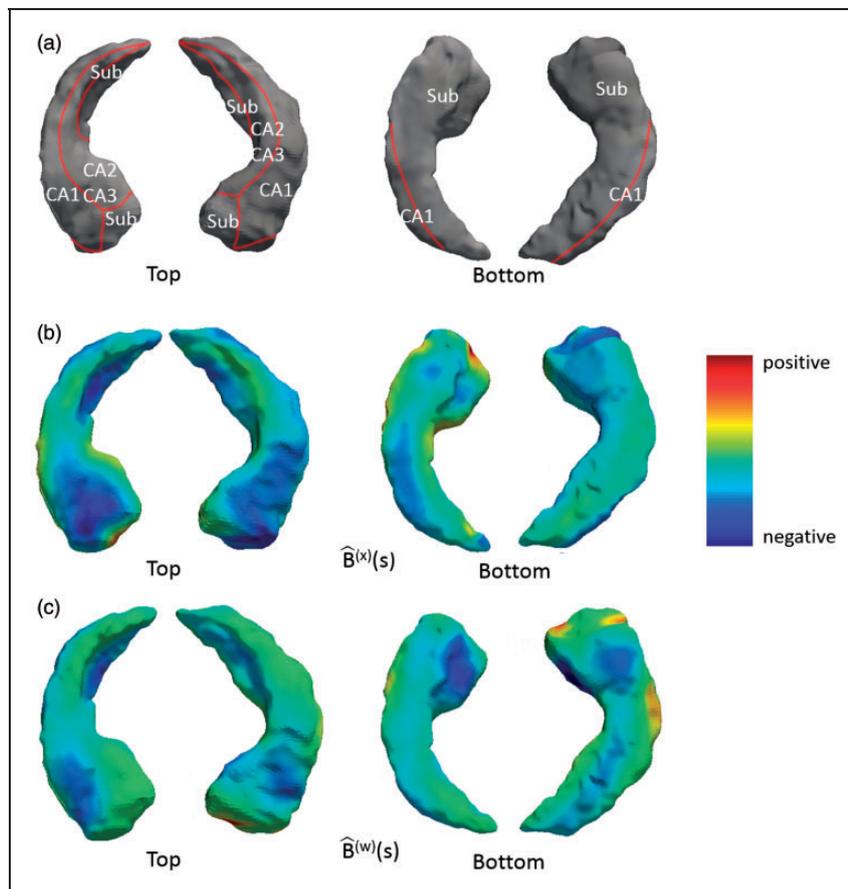

Figure 2. (a) Hippocampal subfields mapped onto the hippocampal surfaces; (b) the estimated coefficient function $\hat{B}^{(x)}(s)$ from the longitudinal submodel; and (c) the estimated coefficient function $\hat{B}^{(w)}(s)$ from the survival submodel, mapped onto the hippocampal surfaces.

$\hat{B}^{(x)}(s)$ of the functional predictor *HRD* in the longitudinal submodel. Blue colors denote negative values of $\hat{B}^{(x)}(s)$ in the regions. It suggests that the decrease of HRD (i.e. hippocampal atrophy) in the blue regions is associated with increasing ADAS-Cog 11 score and deteriorating cognitive functions. Panel (c) displays the coefficient function $\hat{B}^{(w)}(s)$ of *HRD* in the survival submodel. The negative values of $\hat{B}^{(w)}(s)$ in the blue region indicate
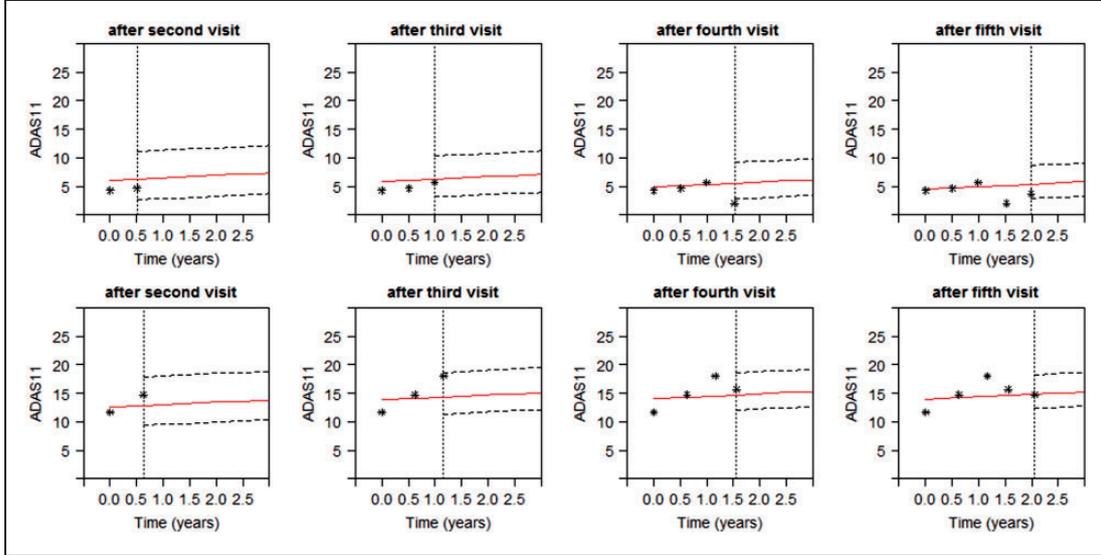
Figure 3. Predicted ADAS-Cog 11 for Patient A (upper panels) and Patient B (lower panels). Solid line is predicted longitudinal trajectories. Dashed lines construct a 95% pointwise uncertainty band. The dotted vertical line represents the time of prediction t. ADAS-Cog: Alzheimer Disease Assessment Scale-Cognitive.

that the thinner of the area, the higher of the risk of progressing from MCI to AD. Most blue regions in Panels (b) and (c) are located in the CA1 subfield and subiculum (Sub) subfield displayed in Panel (a), suggesting that regional radial atrophy in these subfields may be a good predictor of AD progression among MCI patients.

To illustrate the personalized dynamic predictions, we select two target patients as validation data, and predict their future health outcome and event-free probability based on $FJM3$ estimates using the remaining data as training set. Patient A has a baseline age of 73, no $APOE\text{-}\varepsilon$, as compared with a more severe Patient B, 69 years old at baseline, and with $APOE\text{-}\varepsilon$. Figure 3 demonstrates how the predicted ADAS-Cog 11 scores are updated over time for the two patients. From left to right on Figure 3, by using more follow-up data, predictions are closer to the true observed values and the 95% uncertainty band is narrower. It also suggests that Patient A (upper panels) has lower and more stable predicted ADAS-Cog 11 scores (better cognitive function) than patient B (lower panels). Figure 4 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 considerably 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 Bayesian FJM models, and investigate the predictive and calibration performance of the survival probability using the FJM models. We generate 200 datasets with sample size $I = 700$ subjects and each subject has $J_i = 5$ measurements at time 0, 3, 6, 9, and 12. The simulated data structure is similar to the motivating ADNI study, and we include one functional predictor in both longitudinal and survival submodels. The longitudinal submodel is

$$y_i(t_{ij}) = m_i(t_{ij}) + \varepsilon_{ij}$$
$$m_i(t_{ij}) = \beta_0 + \beta_1 t_{ij} + \int_S g_i^{(x)}(s) B^{(x)}(s) \mathrm{d}s + u_i, \quad s \in [0, 1] = S$$

where $j = 1, \ldots, J_i$, $\varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2 = 1)$, and $u_i \sim N(0, \sigma_u^2 = 2)$. The time-invariant functional predictor is defined on a 1D domain $S = [0, 1]$ as $g_i^{(x)}(s) = d_{i1} + d_{i2}s + \sum_{k=1}^{10} \{v_{is1} \cdot \sin(2\pi ks) + v_{is2} \cdot \cos(2\pi ks)\}$, where $d_{i1} \sim U(0, 5)$, $d_{i2} \sim N(1, 0.2)$, and $v_{is1}, v_{is2} \sim N(0, 1/k^2)$. It is observed on a discrete grid at location $s = c/100$, where $c = 0, \ldots, 100$. The observed functional predictor $g_i^{(x')}(s) = g_i^{(x)}(s) + \epsilon_i(s)$, where the measurement errors
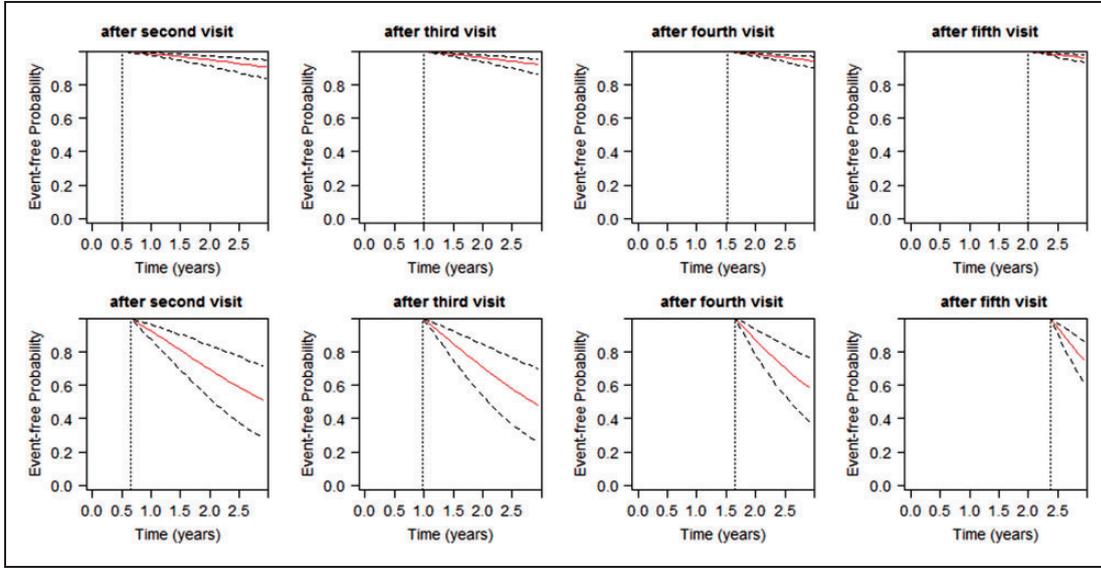
Figure 4. Predicted event-free probability with 95% pointwise uncertainty band for Patient A (upper panels) and Patient B (lower panels). The dotted vertical line represents the time of prediction t.

$\epsilon_i(s) \sim N(0, 0.1)$ across s. The coefficient $\beta_0 = 1.35$, $\beta_1 = 7.0$, and coefficient function $B^{(x)}(s) = -1.5 - \sin(2\pi s) - \cos(2\pi s)$. The survival submodel is

$$h_i(t) = h_0(t) \exp\{\gamma_1 w_1 + \int_S g_i^{(w)}(s) B^{(w)}(s) ds + \alpha m_i(t)\}, \quad s \in [0, 1] = S$$

where the baseline hazard function $h_0(t) = \exp(-10)$, w_1 is simulated from Bernoulli distribution with probability being 0.5, and $\gamma_1 = -2.5$. We have functional predictor $g_i^{(w)}(s) = g_i^{(x)}(s)$ and the observed functional predictor $g_i^{(w')}(s)$ is generated in a similar fashion as $g_i^{(x')}(s)$. The coefficient function $B^{(w)}(s) = \frac{1}{2.5} \Phi(\frac{s-0.6}{0.2^2})$, where $\Phi(\cdot)$ is the standard Normal density function. Censoring time is independently simulated from a uniform distribution to achieve a censoring rate about 30%. Due to censoring, each subject has an average of four repeated measurements. We perform FPCA to the simulated functional predictors $g_i^{(x')}(s)$ and $g_i^{(w')}(s)$, and choose the first 10 FPCs. The coefficient functions $B^{(x)}(s)$ and $B^{(w)}(s)$ are expanded by cubic B-spline basis with $K_B = 10$.

For each of the 200 simulated datasets (denoted by subscript r), we randomly select 500 subjects as the training dataset and set aside the remaining 200 subjects as the validation dataset for prediction. The ability to estimate the true coefficient is assessed by the average mean squared error (AMSE). For the coefficient functions $\hat{B}(\cdot)$, $AMSE(\hat{B}(.)) = \frac{1}{200} \sum_{r=1}^{200} [\hat{B}_r(s) - B(s)]^2$, and for the other parameters, $AMSE(\hat{\beta}_1) = \frac{1}{200} \sum_{r=1}^{200} (\hat{\beta}_{1r} - \beta_1)^2$. 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. Table 3 suggests that the proposed Bayesian FJM performs reasonable well with relatively small AMSE values for both coefficient functions and other parameters, SE being close to SD, and the credible interval coverage probabilities being reasonably close to 95%.

Figure 5 displays the true coefficient functions $B^{(x)}(s)$ and $B^{(w)}(s)$ (red solid lines) and their estimated curves (black solid lines), along with the 95% pointwise credible interval (dashed lines). Both panels suggest that the estimated coefficient functions are reasonably close to the true coefficient functions, with 95% pointwise credible interval always covering the true functions.

For each testing dataset, we predict subject-specific conditional survival probability $\hat{\pi}_k(t'|t)$ for each subject at different time points t and $t' = t + \Delta t$ using the MCMC samples from the fitted model and available measurements up to time t. The predicted time-dependent AUCs are calculated based on the predicted probabilities of all subjects. The true conditional survival probability $\pi_N(t'|t)$ and AUCs are computed using the prespecified parameter values and the generated random effects. We use the dynamic expected BS $\widehat{BS}(t'|t)$ to assess the bias between the predicted and true risks. Table 4 presents the time-dependent AUCs and BSs by averaging the results

Table 3. Parameter estimation in the simulation study.

	Bias	AMSE	SE	SD	CP
For longitudinal data					
$\beta_1 = 7.00$	<0.001	<0.001	0.006	0.007	0.920
$B^{(x)}(s)$		0.014			
$\sigma_\varepsilon^2 = 1$	0.006	0.002	0.038	0.040	0.925
$\sigma_u^2 = 2$	<0.001	0.023	0.144	0.152	0.940
For survival data					
$\gamma_1 = -2.5$	−0.059	0.042	0.198	0.197	0.955
$\alpha = 1.2$	0.021	0.005	0.068	0.069	0.930
$B^{(w)}(s)$		0.003			

AMSE: average mean squared error; SE: standard error; SD: standard deviation; CP: coverage probability.
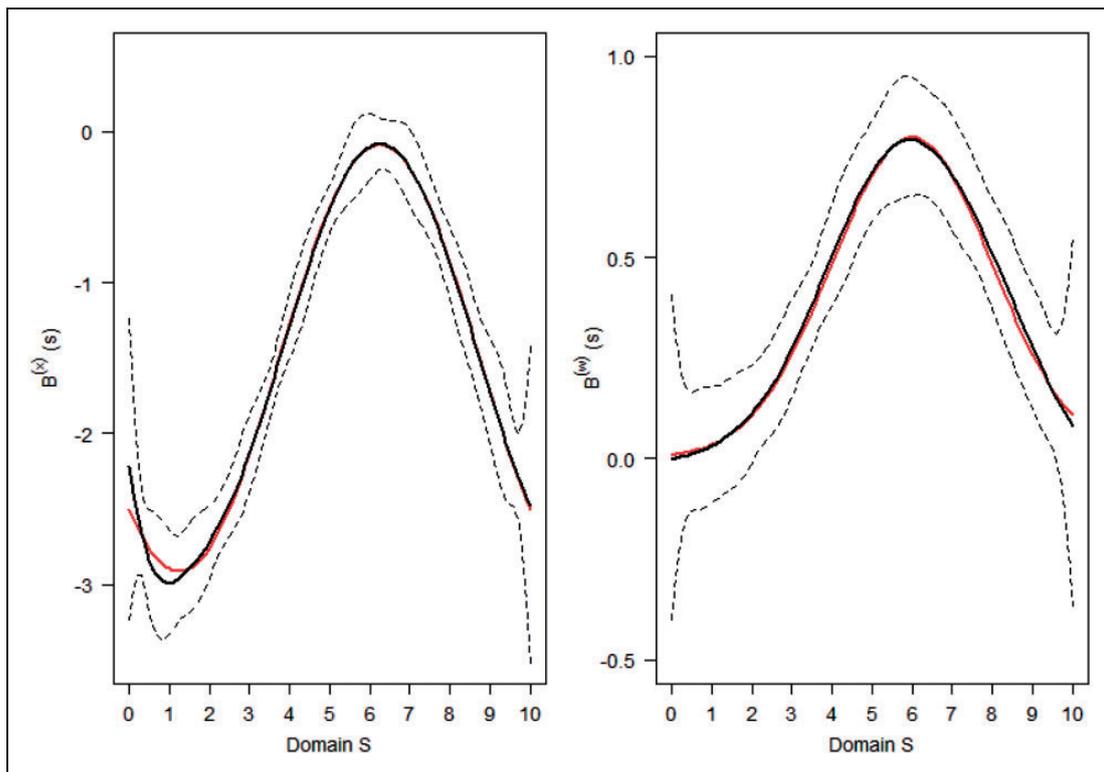
Figure 5. The FJM's estimates of coefficient functions in simulation study. The true coefficient functions (red solid line) $B^{(x)}(s)$ (left panel) and $B^{(w)}(s)$ (right panel) and their estimates (black solid line), along with 95% pointwise credible interval (dashed lines). FJM: functional joint model.

from 200 simulated datasets. The predict AUCs are only slightly smaller than the true AUCs, suggesting good prediction performance of the FJM in terms of validation. The small values in BSs indicate the excellent calibration performance of the FJM.

6 Discussion

The methodology introduced in this article is motivated by many current studies where functional data are collected. The ADNI is just one example of such studies. However, both theoretical and computational complexity in modeling functional data restrict researchers to focused on scalar measures, e.g. volumes of a few brain regions. Without careful analysis of the functional data, pace to understand diseases and population's health condition can be dramatically slowed down. Moreover, some high dimensional data, such as genetic variant

Table 4. AUCs and BSs for the simulation study.

Δt	t	True $AUC_t^{\Delta t}$	Predicted $AUC_t^{\Delta t}$	$\widehat{BS}(t'|t)$
3	3	0.856	0.853	0.015
	6	0.948	0.941	0.024
	9	0.960	0.957	0.035
6	3	0.935	0.929	0.033
	6	0.959	0.957	0.044
	9	0.970	0.959	0.047

AUC: Area under the ROC curve; BS: Brier score.

profiles defined along chromosomes or genomic regions, are commonly treated as functional data. To this end, functional data analysis methods are becoming increasingly important.

In this article, we first propose a Bayesian FJM to account for functional predictors in both longitudinal and survival submodels within the framework of joint modeling. We use the FPCA to approximate the functional predictor, and expand its corresponding coefficient function using penalized B-spline approach. We then develop the process of making personalized dynamic prediction of future outcome trajectories and risks of event using both scalar and functional predictors. Simulation indicates that the proposed Bayesian FJM yields accurate inference and prediction. Being applied to the motivating ADNI study, the FJM can efficiently estimate the association between the trajectory of cognitive functions measured by the ADAS-Cog 11 score and time to AD diagnosis, while accounting for the functional predictor (HRD) and other scalar covariates. We have identified the regional radial atrophy in the CA1 subfield and the subiculum subfield as a good predictor of AD progression among patients with MCI. More importantly, the proposed Bayesian FJM can utilize the functional predictors to make correct predictions for new subjects. The inclusion of functional predictor HRD into both the longitudinal and survival submodels improves overall model fitting and predictive performance. When new measurements are available, the predictions can be updated with improved accuracy and efficiency. Thus, earlier diagnosis can be made to subjects with high predicted risk of deterioration, and intervention can be planned in a timely manner to delaying the manifestation of AD in prodromal AD patients. In addition, although we only use HRD in ADNI data analysis, the proposed Bayesian FJM can readily include multiple brain regions, and even genotype profiles, as functional predictors to provide even precise prediction of Alzheimer's disease progression.

There are some limitations we will address in the future. First, we exclude from analysis the subjects (45 out of 400) with missing data in baseline covariates of interest. Most of these subjects either do not have baseline image data measured or do not have image data in the archive due to the technical difficulty in image data collection and storage. Although the well-established imputation methods, e.g. multiple imputation, can be used for missing data in scalar predictors, it is challenging in both statistical theory and computation to handle missing functional data. We will extend our Bayesian FJM to account for missing data in both scalar and functional covariates in the future. Second, we use time-invariant functional data as predictor in this article. It would be of scientific interest to extend the proposed Bayesian FJM to accommodate longitudinal functional data. We can decompose the longitudinal functional predictor using longitudinal functional principal component analysis (LFPCA)[45] to account for its longitudinal data structure. Alternatively, we can treat the longitudinal functional data as a functional response variable in the longitudinal submodel (function-on-scalar regression problem) and it can be incorporated in the survival submodel as a time-dependent functional predictor. This model can investigate how the longitudinal functional variable directly impacts the risk for an event.

Moreover, in model (2), we implicitly assume that the risk for an event at time t depends on the unobserved true value of the longitudinal outcome at the same time point. We can investigate different functional forms for the association structure between the longitudinal outcome and the risk for an event.[46,47] For example, both the unobserved true value of $m_i(t_{ij})$ in model (1) and its time-dependent slope $m_i'(t_{ij})$ or even its cumulative effect $\int_0^t m_i(s)ds$ can be included in model (2). The proposed Bayesian FJM can readily accommodate various functional forms with minor modification to the sample codes. Furthermore, the Bayesian procedure we develop is shown to have good inferential properties in simulation studies. As sample size and the number of parameters associated with B-spline grow, computation time may increase dramatically. To this end, paralleling MCMC[48] and variational approximations[49] may address the computational concerns and result in good estimations of model components. We would like to investigate these issues in our future research.

Acknowledgements

Declaration of conflicting interests

Funding

References

1. Faucett CL and Thomas DC. Simultaneously modelling censored survival data and repeatedly measured covariates: a Gibbs sampling approach. *Stat Med* 1996; **15**: 1663–1685.
2. Wulfsohn MS and Tsiatis AA. A joint model for survival and longitudinal data measured with error. *Biometrics* 1997; **53**: 330–339.
3. Rizopoulos D. Dynamic predictions and prospective accuracy in joint models for longitudinal and time-to-event data. *Biometrics* 2011; **67**: 819–829.
4. Taylor JM, Park Y, Ankerst DP, et al. Real-time individual predictions of prostate cancer recurrence using joint models. *Biometrics* 2013; **69**: 206–213.
5. Ramsay JO and Dalzell CJ. Some tools for functional data analysis. *J Roy Stat Soc Ser B (Methodol)* 1991; **53**: 539–572.
6. Wang JL, Chiou JM and Müller HG. Functional data analysis. *Annu Rev Stat Appl* 2016; **3**: 257–295.
7. Morris JS. Functional regression. *Annu Rev Stat Appl* 2015; **2**: 321–359.
8. Cardot H, Ferraty F and Sarda P. Spline estimators for the functional linear model. *Stat Sin* 2003; **13**: 571–591.
9. James GM. Generalized linear models with functional predictors. *J Roy Stat Soc: Ser B (Stat Methodol)* 2002; **64**: 411–432.
10. Müller HG and Stadtmüller U. Generalized functional linear models. *Ann Stat* 2005; **33**: 774–805.
11. Reiss PT and Ogden RT. Functional principal component regression and functional partial least squares. *J Am Stat Assoc* 2007; **102**: 984–996.
12. James GM, Wang J and Zhu J. Functional linear regression that's interpretable. *Ann Stat* 2009; **37**: 2083–2108.
13. Goldsmith J, Bobb J, Crainiceanu CM, et al. Penalized functional regression. *J Comput Graph Stat* 2012; **20**: 830–851.
14. Goldsmith J, Crainiceanu CM, Caffo B, et al. Longitudinal penalized functional regression for cognitive outcomes on neuronal tract measurements. *J Roy Stat Soc: Ser C (Appl Stat)* 2012; **61**: 453–469.
15. Gertheiss J, Goldsmith J, Crainiceanu C, et al. Longitudinal scalar-on-functions regression with application to tractography data. *Biostatistics* 2013; **14**: 447–461.
16. Gellar JE, Colantuoni E, Needham DM, et al. Cox regression models with functional covariates for survival data. *Stat Model* 2015; **15**: 256–278.
17. Lee E, Zhu H, Kong D, et al. BFLCRM: A Bayesian functional linear Cox regression model for predicting time to conversion to Alzheimer's disease. *Ann Appl Stat* 2015; **9**: 2153–2178.
18. Weiner MW, Veitch DP, Aisen PS, et al. 2014 Update of the Alzheimer's Disease Neuroimaging Initiative: A review of papers published since its inception. *Alzheimers Dement* 2015; **11**: e1–e120.
19. Li K, Chan W, Doody RS, et al. Prediction of conversion to Alzheimers disease with longitudinal measures and time-to-event data. *J Alzheimers Dis* 2017; **58**: 361–371.
20. Cleveland WS. Robust locally weighted regression and smoothing scatterplots. *J Am Stat Assoc* 1979; **74**: 829–836.
21. Diggle P and Kenward MG. Informative drop-out in longitudinal data analysis. *J Roy Stat Soc Ser C (Appl Stat)* 1994; **43**: 49–93.
22. Lo RY, Hubbard AE, Shaw LM, et al. Longitudinal change of biomarkers in cognitive decline. *Archiv Neurol* 2011; **68**: 1257–1266.
23. Zhang D, Shen D, Initiative ADN, et al. Predicting future clinical changes of MCI patients using longitudinal and multimodal biomarkers. *PLOS One* 2012; **7**: e33182.

24. Du AT. Magnetic resonance imaging of the entorhinal cortex and hippocampus in mild cognitive impairment and Alzheimer's disease. *J Neurol Neurosurg Psychiatry* 2001; **71**: 441–447.
25. Cui Y, Liu B, Luo S, et al. Identification of conversion from mild cognitive impairment to Alzheimer's disease using multivariate predictors. *PLOS One* 2011; **6**: e21896.
26. Li S, Okonkwo O, Albert M, et al. Variation in variables that predict progression from MCI to AD dementia over duration of follow-up. *Am J Alzheimers Dis* 2013; **2**: 12–28.
27. Apostolova LG, Mosconi L, Thompson PM, et al. Subregional hippocampal atrophy predicts Alzheimer's dementia in the cognitively normal. *Neurobiol Aging* 2010; **31**: 1077–1088.
28. Qiu A, Fennema-Notestine C, Dale AM, et al. Regional shape abnormalities in mild cognitive impairment and Alzheimer's disease. *NeuroImage* 2009; **45**: 656–661.
29. Ruppert D. Selecting the number of knots for penalized splines. *J Comput Graph Stat* 2002; **11**: 735–757.
30. Staniswalis JG and Lee JJ. Nonparametric regression analysis of longitudinal data. *J Am Stat Assoc* 1998; **93**: 1403–1418.
31. Yao F, Müller HG, Clifford AJ, et al. Shrinkage estimation for functional principal component scores with application to the population kinetics of plasma folate. *Biometrics* 2003; **59**: 676–685.
32. Stefan Lang AB. Bayesian P-splines. *J Comput Graph Stat* 2004; **13**: 183–212.
33. Hoffman MD and Gelman A. The No-U-turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo. *J Mach Learn Res* 2014; **15**: 1593–1623.
34. Gelman A, Carlin JB, Stern HS, et al. *Bayesian data analysis*. Boca Raton, FL: CRC Press, 2013.
35. Gilks WR, Best NG and Tan KKC. Adaptive rejection Metropolis sampling within Gibbs sampling. *J Roy Stat Soc Ser C (Appl Stat)* 1995; **44**: 455–472.
36. Li L, Greene T and Hu B. 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* 2016; **0**: 0962280216680239.
37. Heagerty PJ, Lumley T and Pepe MS. Time-dependent ROC curves for censored survival data and a diagnostic marker. *Biometrics* 2000; **56**: 337–344.
38. Blanche P, Proust-Lima C, Loubre L, et al. Quantifying and comparing dynamic predictive accuracy of joint models for longitudinal marker and time-to-event in presence of censoring and competing risks. *Biometrics* 2015; **71**: 102–113.
39. Sne M, Taylor JMG, Dignam JJ, et al. Individualized dynamic prediction of prostate cancer recurrence with and without the initiation of a second treatment: development and validation. *Stat Methods Med Res* 2016; **25**: 2972–2991.
40. Crainiceanu CM, Reiss PT, Goldsmith J, et al. Refund: Regression with functional data, http://CRAN.R-project.org/package=refund, r package version 0.1-6 (2013).
41. Peng J and Paul D. A geometric approach to maximum likelihood estimation of the functional principal components from sparse longitudinal data. *J Comput Graph Stat* 2009; **18**: 995–1015.
42. Risacher SL, Saykin AJ, Wes JD, et al. Baseline MRI predictors of conversion from MCI to probable AD in the ADNI cohort. *Curr Alzheimer Res* 2009; **6**: 347–361.
43. Corder E, Saunders A, Strittmatter W, et al. Gene dose of apolipoprotein E type 4 allele and the risk of Alzheimer's disease in late onset families. *Science* 1993; **261**: 921–923.
44. Apostolova LG, Dutton RA, Dinov ID, et al. Conversion of mild cognitive impairment to Alzheimer disease predicted by hippocampal atrophy maps. *Archiv Neurol* 2006; **63**: 693–699.
45. Greven S, Crainiceanu C, Caffo B, et al. Longitudinal functional principal component analysis. *Electron J Stat* 2010; **4**: 1022–1054.
46. Rizopoulos D, Hatfield LA, Carlin BP, et al. Combining dynamic predictions from joint models for longitudinal and time-to-event data using Bayesian model averaging. *J Am Stat Assoc* 2014; **109**: 1385–1397.
47. Yang L, Yu M and Gao S. Prediction of coronary artery disease risk based on multiple longitudinal biomarkers. *Stat Med* 2016; **35**: 1299–1314.
48. Wang X, Guo F, Heller KA, et al. Parallelizing MCMC with random partition trees. In: Cortes C, Lawrence ND and Lee DD (eds) *Advances in neural information processing systems 28*. Red Hook, NY: Curran Associates, Inc., 2015, pp.451–459.
49. Ormerod JT and Wand MP. Explaining variational approximations. *Am Stat* 2010; **64**: 140–153.