

Functional joint model for longitudinal and time-to-event data: an application to Alzheimer's disease

Kan Li and Sheng Luo*[†] 

Functional data are increasingly collected in public health and medical studies to better understand many complex diseases. Besides the functional data, other clinical measures are often collected repeatedly. Investigating the association between these longitudinal data and time to a survival event is of great interest to these studies. In this article, we develop a functional joint model (FJM) to account for functional predictors in both longitudinal and survival submodels in the joint modeling framework. The parameters of FJM are estimated in a maximum likelihood framework via expectation maximization algorithm. The proposed FJM provides a flexible framework to incorporate many features both in joint modeling of longitudinal and survival data and in functional data analysis. The FJM is evaluated by a simulation study and is applied to the Alzheimer's Disease Neuroimaging Initiative study, a motivating clinical study testing whether serial brain imaging, clinical, and neuropsychological assessments can be combined to measure the progression of Alzheimer's disease. Copyright © 2017 John Wiley & Sons, Ltd.

Keywords: ADNI study; functional data analysis; longitudinal and time-to-event data; functional principal component analysis.

1. Introduction

Modern technologies are currently producing increasingly large, complex, and high-dimensional data in medical research. One such type of data is functional data, whose units of observation are functions defined on certain continuous domains (e.g., time, space, or both) but sampled on discrete grids. These functions may be defined on a one-dimensional (1D) Euclidean domain, such as growth curve data, heart rate monitor data, and electroencephalogram data. A growing volume of functional data is also collected on higher dimensional domains such as magnetic resonance imaging (MRI), positron emission tomography, and functional MRI. Moreover, prospective cohort studies and clinical trials investigating neurodegenerative diseases such as Alzheimer's disease (AD) and Huntington's disease often collect repeated measurements of clinical variables, event history, and functional data, which induce complex correlation structures among the observations. The emergence of new data types and structures not only brings rich source of information but also poses new challenges in methodology development of functional data analysis (FDA).

Function regression, especially functional predictor regression (scalar-on-function regression, models the relationship between a scalar outcome and functional predictors), is an active area of FDA in the past 10 years. It was first introduced by Ramsay and Dalzell [1] and built up by Ramsay and Silverman [2]. There is a rich literature in functional predictor regression [e.g., 3–10], while most existing work deals only with cross-sectional data. Goldsmith *et al.* [11] first extended the penalized functional regression approach [10] to handle longitudinal measurements in both the response variable and functional predictors by incorporating scalar random effects. Gertheiss *et al.* [12] extended the functional principal component (FPC) regression to longitudinal functional data and allowed for different effects of subject-specific curves. More recently, Gellar *et al.* [13] extended the Cox proportional hazards model to

Department of Biostatistics, The University of Texas Health Science Center at Houston, Houston 77030, TX, U.S.A.

*Correspondence to: Sheng Luo, Department of Biostatistics, The University of Texas Health Science Center at Houston, 1200 Pressler St, Houston 77030, TX, U.S.A.

[†]E-mail: sheng.t.luo@uth.tmc.edu

incorporate functional predictors and estimated the parameters via penalized partial likelihood approach. Lee *et al.* [14] developed a Bayesian functional Cox regression model with both functional and scalar covariates but used different regularization methods. To the best of our knowledge, there is no functional regression modeling attempts to simultaneously analyze the longitudinal measurements and time-to-event data under the joint modeling framework.

Joint models of longitudinal and time-to-event data were proposed by Faucett and Thomas [15] and Wulfsohn and Tsiatis [16]. 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. Tsiatis and Davidian [17] and Proust-Lima *et al.* [18] gave excellent review of joint modeling research. However, current state-of-the-art joint models do not incorporate functional predictors.

The major objective of this article is to incorporate 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 covariates and functional predictors. The rest of the article is organized as follows. In Section 2, we describe a motivating clinical study and the data structure. In Section 3, we discuss the joint longitudinal-survival model with functional predictors and the estimation procedure. In Section 4, we apply the proposed FJM to the motivating Alzheimer's Disease Neuroimaging Initiative (ADNI) study. In Section 5, we conduct a simulation study to examine the performance of the proposed 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 ADNI study was launched in year 2003 with the primary goal of testing whether serial MRI, positron emission tomography, cerebrospinal fluid 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>.

Because MCI is commonly considered as a transitional stage between normal cognition and AD, numerous recent studies are to assess various neuroimaging techniques and clinical markers to predict AD diagnosis among MCI patients [e.g., 19]. To this end, we select 355 MCI patients in ADNI-1 study without missing data in covariates of interests and consider AD diagnosis to be the survival event of interest. Participants were followed up for a mean of 3.2 years (standard deviation (SD) 2.6; range 0.4–9.3) before AD diagnosis or censoring. Among the 355 MCI patients, 180 patients were diagnosed with AD, and 175 had stable MCI over a mean follow-up period of 2.3 and 4.2 years, respectively.

Alzheimer's Disease Assessment Scale-Cognitive (ADAS-Cog) score (with 11 items, referred to as ADAS-Cog 11) measures cognition functions, and it is usually reported as a composite score of the 11 items, and it ranges from 0 to 70 (or 85), with a higher score indicating poor cognitive function. The longitudinal ADAS-Cog score was reported to be the strongest longitudinal biomarker predicting time from MCI-to-AD diagnosis [20]. Figure 1 displays the lowest smoothing curve [21] of ADAS-Cog 11 scores over time for the MCI patients, with follow-up time less than 3 years (203 patients, solid line), 3–6 years (82 patients, dotted line), and more than 3 years (70 patients, dashed line), in addition to 95% pointwise confidence intervals (shaded regions). Figure 1 suggests that the ADAS-Cog 11 scores of patients in all three groups increase with time (deteriorating cognitive functions). Moreover, 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. This phenomenon manifests strong correlation between the longitudinal ADAS-Cog 11 values and the time to AD diagnosis. Such a dependent terminal event time is often referred to as dependent censoring or informative censoring in the literature of joint modeling [22]. However, many studies [23, 24] designed to explore longitudinal measures for predicting future cognitive functions of MCI patients fail to account for such informative censoring, leading to biased inference.

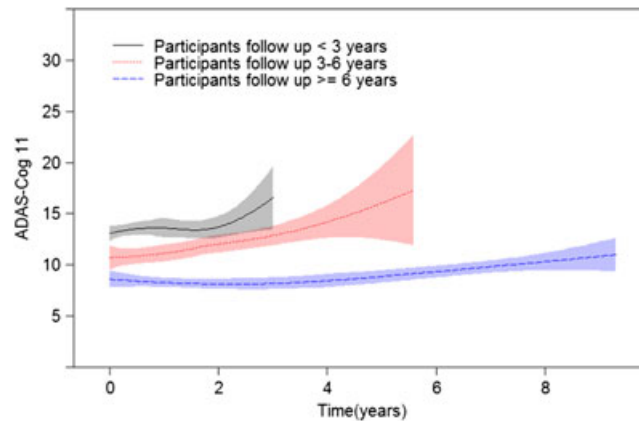


Figure 1. Smoothing curves of variable Alzheimer’s Disease Assessment Scale-Cognitive 11 (ADAS-Cog 11) over time for mild cognitive impairment patients with follow-up time less than 3 years (203 patients, solid line), 3–6 years (82 patients, dotted line), and more than 6 years (70 patients, dashed line).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

Moreover, degree of atrophy within the medial temporal lobe structures, especially within the hippocampus, was reported to be associated with the AD progression. Most of the current analysis was based on volumetric brain MRI data. 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 [25]. In the preliminary analysis, we fit regular joint models with baseline MRI measure including bilateral hippocampal volume as a scalar covariate. The results (detailed in the Supporting Information) suggest that hippocampal volume has strong association with both cognitive function decline and time to AD diagnosis. Furthermore, some researches [26–29] demonstrated that the surface-based morphology analysis offers more advantages because these methods study patterns of hippocampal subfield atrophy and produce detailed pointwise correlation between atrophy and cognitive functions.

In this article, we propose an FJM to examine the association between the 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 each bilateral hippocampal surface point (referred to as vertex), which measures the distance from the medial core to each vertex on the surface and represents the hippocampal thickness. For image processing, we adopt a surface fluid registration package [30]. The left and right hippocampal surfaces are first conformally mapped to a two-dimensional (2D) rectangle plane, in the form of matrix, to form a feature image of the surface. After registering each feature image to a common template and calculating HRD for all vertexes, the 2D image matrices are vectorized into 1D image vectors. The spatial information and image smoothing have been accounted for in part in the image preprocessing steps as described in Shi *et al.* [30]. Our focus here is to propose methodology that is applicable to a wide variety of imaging and non-imaging functional data. To this end, the corresponding HRDs of the vertexes are represented as a 1D functional data defined on domain S . Each point in the image vector retains a one-to-one relationship to the original vertex on the surface, which allows us to back-transform any functions defined on domain S to the hippocampal surfaces. The hippocampus image processing procedure is detailed in the Supporting Information.

3. Functional joint model

We introduce a general FJM framework for a longitudinal process and a time-to-event process with time-invariant functional predictors. For the ease of illustration, we only incorporate a single time-invariant (baseline) functional predictor in both the longitudinal and survival submodels, while the FJM can readily accommodate multiple functional predictors.

3.1. Functional joint model framework

For each subject i ($i = 1, \dots, I$) at visit j ($j = 1, \dots, J_i$), we observe data $\{y_{ij}, \mathbf{x}_{ij}, g_i^{(x)}(s)\}$, where $y_{ij} = y_i(t_{ij})$ is a scalar outcome recorded at time t_{ij} from the study onset. Vector \mathbf{x}_{ij} is a p -dimensional covariate vector. Function $g_i^{(x)}(s)$ is a time-invariant functional predictor defined over a 1D domain $s \in [0, S_{\max}] = S$.

The longitudinal submodel is

$$y_i(t_{ij}) = m_i(t_{ij}) + \varepsilon_{ij}, \quad \text{where } m_i(t_{ij}) = \beta_0 + \mathbf{x}_{ij}^\top \boldsymbol{\beta} + \int_S g_i^{(x)}(s) B^{(x)}(s) ds + \mathbf{z}_{ij}^\top \mathbf{u}_i \quad (1)$$

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 \mathbf{z}_{ij} is a q -dimensional covariate corresponding to random effects \mathbf{u}_i , which are assumed to have $\mathbf{u}_i \sim N(0, \boldsymbol{\Sigma}_u)$ to account for the within-subject correlation. The measurement error $\varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2)$ is independent from \mathbf{u}_i .

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

$$h(t) = h_0(t) \exp\{\mathbf{w}_i^\top \boldsymbol{\gamma} + \int_S g_i^{(w)}(s) B^{(w)}(s) ds + \alpha m_i(t)\}, \quad (2)$$

where $h_0(t)$ is the baseline hazard function and \mathbf{w}_i is a vector of time-independent covariates with regression coefficient vector $\boldsymbol{\gamma}$. 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. Models (1) and (2) consist of the FJM framework.

3.2. Functional regression

To build the functional regression model, we follow the general strategy of FPC regression (e.g., [4, 8]). 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 $\Sigma^{(x)}(s, s') = \text{cov}\{g_i^{(x)}(s), g_i^{(x)}(s')\}$ 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 $\lambda_1^{(x)} \geq \lambda_2^{(x)} \geq \dots \geq 0$ are non-increasing eigenvalues and $\phi_l^{(x)}(s)$ 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 FPC scores $\xi_{il}^{(x)} = \int_S \{g_i^{(x)}(s) - \mu^{(x)}(s)\} \phi_l^{(x)}(s) ds$ 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 $g_i^{(x)}(s) \approx \mu^{(x)}(s) + \sum_{l=1}^{K_x} \xi_{il}^{(x)} \phi_l^{(x)}(s)$. The number of components K_x can be determined using the proportion of explained variance. 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 proportion of explained variance, for example, $L = 80\%$, 90% , or 95% .

In the second step, we use the first K_x eigenfunctions as the basis functions to expand the coefficient function $B^{(x)}(s)$ in model (1) as $B^{(x)}(s) = \sum_{l=1}^{K_x} B_l^{(x)} \phi_l^{(x)}(s)$, where the coefficient $B_l^{(x)} = \int_S B^{(x)}(s) \phi_l^{(x)}(s) ds$. We let vector of FPC scores $\boldsymbol{\xi}_i^{(x)} = [\xi_{i1}^{(x)}, \dots, \xi_{iK_x}^{(x)}]^\top$, vector of eigenfunctions $\boldsymbol{\phi}^{(x)}(s) = [\phi_1^{(x)}(s), \dots, \phi_{K_x}^{(x)}(s)]^\top$, and vector of coefficients $\mathbf{B}^{(x)} = [B_1^{(x)}, \dots, B_{K_x}^{(x)}]^\top$. Then, we have

$$\begin{aligned} \int_S g_i^{(x)}(s) B^{(x)}(s) ds &\approx \int_S \mu^{(x)}(s) B^{(x)}(s) ds + \int_S (\boldsymbol{\xi}_i^{(x)})^\top \boldsymbol{\phi}^{(x)}(s) \boldsymbol{\phi}^{(x)}(s)^\top \mathbf{B}^{(x)} ds \\ &= \int_S \mu^{(x)}(s) B^{(x)}(s) ds + (\boldsymbol{\xi}_i^{(x)})^\top \int_S \boldsymbol{\phi}^{(x)}(s) \boldsymbol{\phi}^{(x)}(s)^\top ds \mathbf{B}^{(x)} = \int_S \mu^{(x)}(s) B^{(x)}(s) ds + (\boldsymbol{\xi}_i^{(x)})^\top \mathbf{B}^{(x)}. \end{aligned}$$

Note that $\int_S \boldsymbol{\phi}^{(x)}(s) \boldsymbol{\phi}^{(x)}(s)^\top ds = \mathbf{I}$, where \mathbf{I} is an identity matrix, because the eigenfunctions $\phi_l^{(x)}(s)$ are orthonormal.

Similarly, the functional predictor $g_i^{(w)}(s)$ 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 thus $\int_S g_i^{(w)}(s)B^{(w)}(s)ds \approx \int_S \mu^{(w)}(s)B^{(w)}(s)ds + (\xi_i^{(w)})^\top B^{(w)}$, where $\mu^{(w)}(s)$, $\xi_{il}^{(w)}$, $\phi_l^{(w)}(s)$, and $B^{(w)}$ have the same meanings as $\mu^{(x)}(s)$, $\xi_{il}^{(x)}$, $\phi_l^{(x)}(s)$, and $B^{(x)}$, respectively. Thus, the FJM based on the FPC scores is

$$y_i(t) = m_i(t) + \varepsilon_{ij}, \text{ where } m_i(t) = \beta'_0 + \mathbf{x}_{ij}^\top \boldsymbol{\beta} + (\xi_i^{(x)})^\top \mathbf{B}^{(x)} + \mathbf{z}_{ij}^\top \mathbf{u}_i, \tag{3}$$

$$\text{and } h(t) = h_0^*(t) \exp\{\mathbf{w}_i^\top \boldsymbol{\gamma} + (\xi_i^{(w)})^\top \mathbf{B}^{(w)} + \alpha m_i(t)\}, \tag{4}$$

where $\beta'_0 = \beta_0 + \int_S \mu^{(x)}(s)B^{(x)}(s)ds$ and $h_0^*(t) = h_0(t) \exp\{\int_S \mu^{(w)}(s)B^{(w)}(s)ds\}$. 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 model for the longitudinal scalar response variable and a Cox model for the survival outcome, respectively, in a standard joint model framework [16]. The 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 $\boldsymbol{\theta} = [\boldsymbol{\beta}^\top, (\mathbf{B}^{(x)})^\top, \text{vech}(\boldsymbol{\Sigma}_u)^\top, \boldsymbol{\gamma}^\top, (\mathbf{B}^{(w)})^\top, \alpha, \sigma_\varepsilon^2, \boldsymbol{\theta}_{h_0^*}^\top]^\top$ be the parameter vector, where $\text{vech}(\boldsymbol{\Sigma}_u)$ is the vector being formed by vectorizing the lower triangular part of covariance matrix $\boldsymbol{\Sigma}_u$ and vector $\boldsymbol{\theta}_{h_0^*}$ denotes the parameters in the baseline hazard function $h_0^*(\cdot)$. The conditional likelihood from the longitudinal data $\mathbf{y}_i = [y_{i1}, \dots, y_{iJ}]^\top$ is

$$p(\mathbf{y}_i | \boldsymbol{\theta}, \mathbf{u}_i) = (2\pi\sigma_\varepsilon^2)^{-J/2} \exp\left\{-\frac{1}{2\sigma_\varepsilon^2} \sum_{j=1}^J \left[y_{ij} - (\beta'_0 + \mathbf{x}_{ij}^\top \boldsymbol{\beta} + (\xi_i^{(x)})^\top \mathbf{B}^{(x)} + \mathbf{z}_{ij}^\top \mathbf{u}_i)\right]^2\right\},$$

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

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

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

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

3.3. Estimation and inference

In practice, the functional predictors such as $g_i^{(x)}(s)$ are measured over finite grids in domain \mathcal{S} and often with error, that is, the observed functional predictor $g_i^{(x)}(s) = g_i^{(x)}(s) + \varepsilon_i(s)$, where measurement error $\varepsilon_i(s) \sim N(0, \sigma_\varepsilon^2)$. 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 [31, 32]. Then, the estimated eigenvalues $\hat{\lambda}_l^{(x)}$ and the corresponding estimated eigenfunctions $\hat{\phi}_l^{(x)}(s)$, where $l = 1, \dots, S_{max}$, are calculated based on the decomposition of the smoothed covariance function. Finally, the estimated FPC scores for each subject are calculated as $\hat{\xi}_{il}^{(x)} = \int_S \{g_i^{(x)}(s) - \hat{\mu}^{(x)}(s)\} \hat{\phi}_l^{(x)}(s) ds$, 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 $\hat{\boldsymbol{\phi}}^{(x)}(s) = [\hat{\phi}_1^{(x)}(s), \dots, \hat{\phi}_{K_x}^{(x)}(s)]^\top$ and $\hat{\boldsymbol{\xi}}_i^{(x)} = [\hat{\xi}_{i1}^{(x)}, \dots, \hat{\xi}_{iK_x}^{(x)}]^\top$, respectively.

Maximization of the likelihood function in model (5) with respect to parameter vector $\boldsymbol{\theta}$ can be performed using expectation maximization (EM) algorithm. More details regarding the EM algorithm

implementation are given in the Supporting Information. Based on the estimated coefficient vector $\hat{\mathbf{B}}^{(x)}$, the estimated coefficient function is calculated by $\hat{B}^{(x)}(s) = (\hat{\mathbf{B}}^{(x)})^\top \hat{\boldsymbol{\phi}}^{(x)}(s)$. A pointwise 95% confidence interval for $B^{(x)}(s)$ can be constructed based on D (e.g., $D = 1000$) bootstrap samples [33], for example, at location s , a 95% bootstrap confidence interval for $B^{(x)}(s)$ can be $[\hat{q}_{B,0.025}(s), \hat{q}_{B,0.975}(s)]$, where $\hat{q}_{B,p}(s)$ is the p -quartile of the bootstrap samples $\hat{B}_d^{(x)}(s)$, $d = 1, \dots, D$. Alternatively, a Wald-type confidence interval based on the SD of the bootstrap estimates, $\hat{s}_B(s)$, is given by $\hat{B}^{(x)}(s) \pm 1.96\hat{s}_B(s)$. These two types of bootstrap confidence intervals give very similar results in our simulation study. Similarly, the estimated coefficient function $\hat{B}^{(w)}(s)$ and its confidence interval can be obtained. For visualization in the ADNI study, we can map the coefficient function $\hat{B}^{(x)}(s)$ back to the hippocampal surfaces, because there is a one-to-one relationship between the location s in domain S and the vertex on the hippocampal surfaces.

3.4. Implementation using software

An advantage of the proposed FJM is that its implementation can be performed using available standard software. The first step is to conduct FPC analysis (FPCA) for functional predictors and to estimate the FPC eigenfunctions and scores, using `fPCA.sc` function in the `refund` package [34] or `fPCA.mle` and `fPCA.score` functions in the `FPCA` package [35] in R. In the second step, FPC scores are used as scalar covariates in a standard joint model for longitudinal and survival data, and their coefficients can be estimated using `JM` package [36] in R. The estimated coefficient function can be calculated as weighted sum of FPC eigenfunctions. In addition, we have fitted the proposed FJM via our own code and have obtained the estimation results very close to those from the aforementioned packages. To facilitate easy reading and implementation of the proposed FJM, we provide in the Supporting Information the R codes to conduct FPCA and to fit the FJM using either the `JM` package or our estimation methods based on EM algorithm.

4. Application to the Alzheimer's Disease Neuroimaging Initiative study

We apply the proposed FJM to the motivating ADNI 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-ε4`, 56%), given their potential effects on AD progression [37–39]. To utilize the brain imaging information, we include baseline hippocampal volume (`bHV`) as a scalar covariate and the baseline hippocampal surfaces based on HRD as a functional predictor. We follow the procedure in the Supporting Information to convert the three-dimensional (3D) HRD to a 1D domain denoted by S .

The first model we consider is the regular joint model (refer to as model `JM`, identified in the preliminary analysis in Section 2), which incorporates variable `bHV` in both longitudinal and survival submodels. Additionally, we consider three FJMs, that is, model `FJM1` includes HRD only in the longitudinal submodel and model `FJM2` includes HRD only in the survival submodel, while model `FJM3` includes HRD in both submodels as

$$\begin{aligned} ADAS - Cog_i(t_{ij}) &= m_i(t_{ij}) + \varepsilon_{ij} \\ m_i(t_{ij}) &= \beta_0 + \beta_1 t_{ij} + \beta_2 gender_i + \beta_3 bAge_i + \beta_4 Edu_i + \beta_5 APOE + \beta_6 bHV_i \\ &\quad + \int_S HRD_i(s) B^{(x)}(s) ds + u_{1i} + u_{2i} t_{ij} \\ h(t) &= h_0(t) \exp\{\gamma_1 gender_i + \gamma_2 bAge_i + \gamma_3 Edu_i + \gamma_4 APOE - \varepsilon_4 \\ &\quad + \gamma_5 bHV_i + \int_S HRD_i(s) B^{(w)}(s) ds + am_i(t)\}. \end{aligned}$$

We perform FPCA on HRD and select the first 20 FPCs, which explain 82.6% of the total variance in the hippocampus radial distance data. 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.

Table I displays the values of Akaike information criterion (AIC) from the four candidate models. The `FJM1` and `FJM3` have smaller AIC than model `JM`, suggesting that including HRD as a functional predictor in the longitudinal submodel may improve the model fit. Model `FJM1` is selected as the final model

	JM	FJM1	FJM2	FJM3
AIC	10211	10202	10217	10208

ADNI, Alzheimer's Disease Neuroimaging Initiative.

	Parameters	MLE	SE	<i>p</i>
For longitudinal outcome				
ADAS-Cog 11	Time (years)	1.006	0.064	<0.001
	Female	-0.227	0.375	0.546
	<i>bAge</i>	-0.259	0.177	0.143
	<i>Edu</i> (years)	-0.248	0.051	<0.001
	<i>APOE</i> - ϵ 4	1.090	0.278	<0.001
	<i>bHV</i> (mm ³)	-0.954	0.201	<0.001
For survival process				
MCI to AD	Female	-0.203	0.168	0.227
	<i>bAge</i>	-0.165	0.087	0.059
	<i>Edu</i> (years)	-0.002	0.026	0.951
	<i>APOE</i> - ϵ 4	0.364	0.168	0.030
	<i>bHV</i> (mm ³)	-0.300	0.091	<0.001
	α	0.107	0.015	<0.001

AD, Alzheimer's disease; ADAS-Cog 11, Alzheimer's Disease Assessment Scale-Cognitive 11; ADNI, Alzheimer's Disease Neuroimaging Initiative; APOE, apolipoprotein E; MCI, mild cognitive impairment; MLE, maximum-likelihood estimate; SE, standard error.

because it has the smallest AIC value. This may indicate that after adjustment of hippocampal volume and other covariates, HRD remains an important functional predictor for the cognitive functions manifested by variable ADAS-Cog 11 among the MCI patients. The HRD may not be significantly associated with the time to AD diagnosis, because FJM2 has the largest AIC.

Parameter estimates from model FJM1 are presented in Table II, while the estimated vector of coefficients $\hat{\mathbf{B}}^{(x)}$ (for a vector of 20 FPC scores $\xi_i^{(x)}$ as in model (3)) are presented in Table SIII. ADAS-Cog 11 score increases (deteriorates) as time progresses, that is, an average increase of 1.006 unit (95% CI: [0.881–1.131]) per year for MCI patients. Higher education, lack of APOE- ϵ 4 allele(s), and larger hippocampal volume at baseline are associated with lower (better) ADAS-Cog 11 scores. Moreover, Table SIII suggests that the coefficients for eight FPC scores are significant ($p < 0.05$), indicating that the baseline HRD is associated with the ADAS-Cog 11 score at all visits. In the survival submodel, the presence of APOE- ϵ 4 allele(s) increases the hazard of AD diagnosis by 44% ($\exp(0.364) - 1$, 95% CI: [4–100%]), which is consistent with the literature [40]. Older age and larger hippocampal volume at baseline are associated with lower risk of AD diagnosis. Furthermore, larger ADAS-Cog 11 score increases the risk of AD diagnosis, that is, one unit increase in ADAS-Cog 11 score increases the hazard of AD diagnosis by 11% ($\exp(0.107) - 1$, 95% CI: [8–15%]).

The coefficient function $B^{(x)}(s)$ is estimated via $\hat{B}^{(x)}(s) = \sum_{l=1}^{20} \hat{B}_l^{(x)} \hat{\phi}_l(s)$. For visualization purpose, each point in the 1D domain \mathcal{S} , along with the coefficient function on that point, is mapped back to the corresponding vertex on the hippocampal surfaces (Figure 2). Because of 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.* [41], on the hippocampal surface template. Panel (b) displays the coefficient function $\hat{B}^{(x)}(s)$ 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. Most blue regions in Panel (b) are located in the CA1 subfield and subiculum (Sub) subfield displayed in Panel (a), suggesting

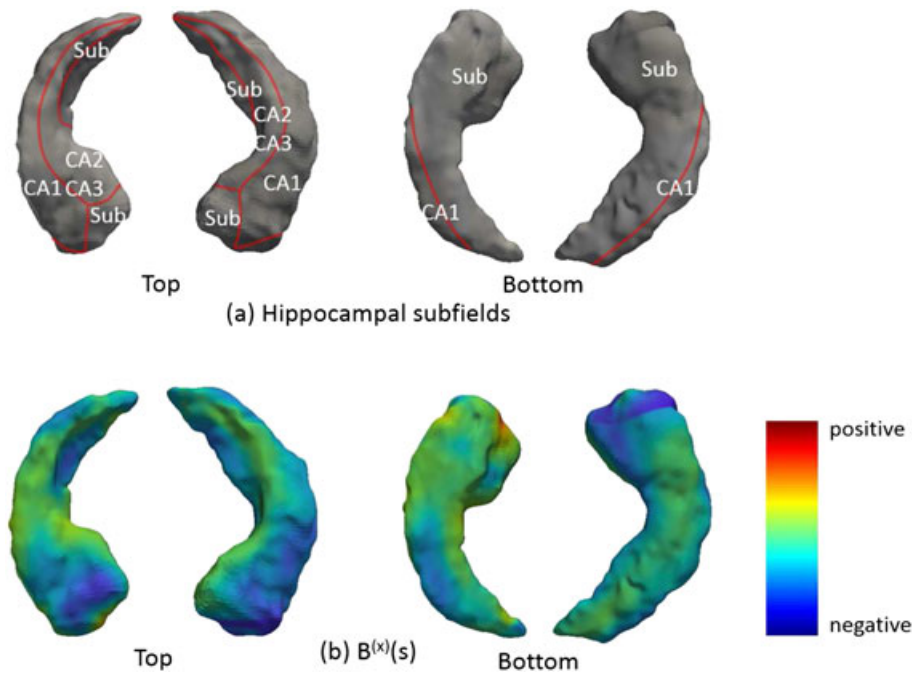


Figure 2. Hippocampal subfields mapped onto the hippocampal surfaces (Panel (a)) and the estimated coefficient function $\hat{B}^{(x)}(s)$ from the longitudinal submodel, mapped onto the hippocampal surfaces (Panel (b)).

that regional radial atrophy in these subfields may be a good predictor of AD progression among MCI patients. The similar point was made in the previous literature [41, 42].

5. Simulation study

In this section, we conduct a simulation study with two settings to evaluate the proposed FJM models. In Setting I, we include one functional predictor in both longitudinal and survival submodels, while in Setting II, we simulate a functional predictor and its coefficient, which are similar to Section 4.

In Setting I, we select $I = 200$ or 500 subjects, and each subject has $J_i = 4$ measurements at time 0, 40, 80, and 120. The longitudinal submodel is

$$\begin{aligned}
 y_i(t_{ij}) &= m_i(t_{ij}) + \varepsilon_{ij}, \\
 m_i(t_{ij}) &= \beta_0 + \beta_1 t_{ij} + \int_S g_i^{(x)}(s) B^{(x)}(s) ds + u_i, \quad s \in [0, 10] = S, \\
 g_i^{(x)}(s) &= u_{i1} + u_{i2} s + \sum_{k=1}^{10} \left\{ v_{is1} \times \sin\left(\frac{\pi k}{5} s\right) + v_{is2} \times \cos\left(\frac{\pi k}{5} s\right) \right\},
 \end{aligned}$$

where $j = 1, \dots, J_i$, $\varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2 = 1)$, $u_i \sim N(0, \sigma_u^2 = 1)$, $u_{i1} \sim U(0, 5)$, $u_{i2} \sim N(1, 0.2)$, and $v_{is1}, v_{is2} \sim N(0, 1/k^2)$. The time-invariant functional predictor is defined on a 1D domain $S = [0, 10]$, and it is observed on a discrete grid at location $s = m/10$, where $m = 0, \dots, 100$. The observed functional predictor $g_i^{(x)}(s) = g_i^{(x)}(s) + \varepsilon_i(s)$, where the measurement errors $\varepsilon_i(s) \sim N(0, 0.1)$ across s . The coefficient function $B^{(x)}(s) = 2 \sin(\pi s/5)$. The survival submodel is

$$h(t) = h_0(t) \exp\left\{ \gamma_1 w_1 + \int_S g_i^{(w)}(s) B^{(w)}(s) ds + a m_i(t) \right\}, \quad s \in [0, 10] = S,$$

where the baseline hazard function $h_0(t) = 0.02$, w_1 is simulated from Bernoulli distribution with probability being 0.5, and $\gamma_1 = -1.75$. 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 $B^{(w)}(s) = 1.2 \sin(\pi s/4)$. Censoring times are independently simulated from a uniform distribution $U(0, c)$, where c

is chosen to achieve a desired censoring rate of 30% or 50%. Because of censoring, each subject has 2 to 3 repeated measurements. We perform FPCA to the simulated functional predictors $g_i^{(x)}(s)$ and $g_i^{(w)}(s)$ and choose the first five FPCs, which explain 95% of the total variance in the original data.

We generate 200 simulated datasets (denoted by subscript r) for each combination of sample sizes ($I = 200$ or 500) and censoring rates (30% or 50%). Table III presents the average mean squared error (AMSE) for the coefficient functions $\hat{B}(\cdot)$ and other parameters as $AMSE(\hat{B}(\cdot)) = \frac{1}{200} \sum_{r=1}^{200} [\hat{B}_r(s) - B(s)]^2$ and $AMSE(\hat{\beta}_1) = \frac{1}{200} \sum_{r=1}^{200} (\hat{\beta}_{1r} - \beta_1)^2$, respectively, in addition to standard error (the square root of the average of the variance), SD (the SD of the maximum likelihood estimations), and coverage probabilities (CP) of 95% confidence intervals. Table III suggests that in Setting I, the proposed FJM performs reasonably well with relatively small AMSE values for both coefficient functions and other parameters, standard error being close to SD, and the credible interval being reasonably close to 95%.

Figure 3 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 confidence bands (shaded regions, constructed using Wald-type confidence intervals based on 1000 bootstrap samples) and the 95% CP. All panels suggest that the estimated coefficient functions from the FJM are reasonably close to the true coefficient functions, with 95% pointwise confidence intervals always covering the true functions. The empirical CP on all regions over the domain S are close to the nominal level of 95%, except the rightmost tail of the coefficient function $B^{(w)}(s)$ in the survival submodel. This may be because that $B^{(w)}(s)$ is not well expanded by the first few principal components calculated from $g_i^{(w)}(s)$ on the rightmost tail. This limitation is further discussed in Section 6.

In Setting II, we keep the same parameters as in Setting I but simulate a functional predictor, which is mapped to a 3D bean surface to resemble the hippocampal surfaces. We evaluate the performance of our FJM when the 3D functional predictor is converted to a 1D vector, as described in the Supporting Information. To do this conversion, we first construct the triangular surface meshes of the bean surface and then conformally map the triangular surface meshes to a rectangle plane. We align the points on the rectangle plane to form a 1D vector but still retain the one-to-one correspondence with points on the original bean surface.

To generate the 1D functional predictor, we use the first 25 orthonormal eigenfunctions $\phi(s) = [\phi_1(s), \dots, \phi_{25}(s)]^T$ and corresponding eigenvalues $\lambda = [\lambda_1, \dots, \lambda_{25}]^T$ that derived from the HRD in

Table III. Parameter estimation in simulation Setting I.

	$n = 200, CR = 0.3$					$n = 500, CR = 0.3$					
	Bias	AMSE	SE	SD	CP	Bias	AMSE	SE	SD	CP	
For longitudinal data											
$\beta_1 = 0.78$	<0.001	<0.001	0.002	0.002	0.970	<0.001	<0.001	0.001	0.002	0.940	
$B^{(x)}(s) = 2 \sin(\pi s/5)$		0.008					0.003				
$\sigma_\epsilon^2 = 1$	<0.001	<0.001	0.030	0.028	0.965	0.007	<0.001	0.019	0.022	0.920	
$\sigma_u^2 = 1$	0.052	0.017	0.130	0.121	0.970	0.005	0.009	0.081	0.092	0.950	
For survival data											
$\gamma_1 = -1.75$	0.077	0.060	0.219	0.233	0.930	0.012	0.027	0.152	0.163	0.940	
$\alpha = 0.29$	0.012	<0.001	0.021	0.022	0.940	0.003	<0.001	0.013	0.016	0.925	
$B^{(w)}(s) = 1.2 \sin(\pi s/4)$		0.023					0.012				
		$n = 200, CR = 0.5$					$n = 500, CR = 0.5$				
For longitudinal data											
$\beta_1 = 0.78$	<0.001	<0.001	0.002	0.002	0.930	<0.001	<0.001	0.001	0.001	0.985	
$B^{(x)}(s) = 2 \sin(\pi s/5)$		0.007					0.003				
$\sigma_\epsilon^2 = 1$	0.002	0.001	0.030	0.030	0.955	<0.001	<0.001	0.019	0.017	0.960	
$\sigma_u^2 = 1$	0.044	0.019	0.130	0.130	0.955	0.016	0.007	0.081	0.083	0.940	
For survival data											
$\gamma_1 = -1.75$	0.115	0.082	0.255	0.262	0.950	0.043	0.021	0.154	0.160	0.950	
$\alpha = 0.29$	0.019	0.001	0.024	0.025	0.920	0.008	<0.001	0.015	0.015	0.915	
$B^{(w)}(s) = 1.2 \sin(\pi s/4)$		0.031					0.010				

AMSE, average mean squared error; CP, coverage probability; CR, censoring rate; SD, standard deviation; SE, standard error.

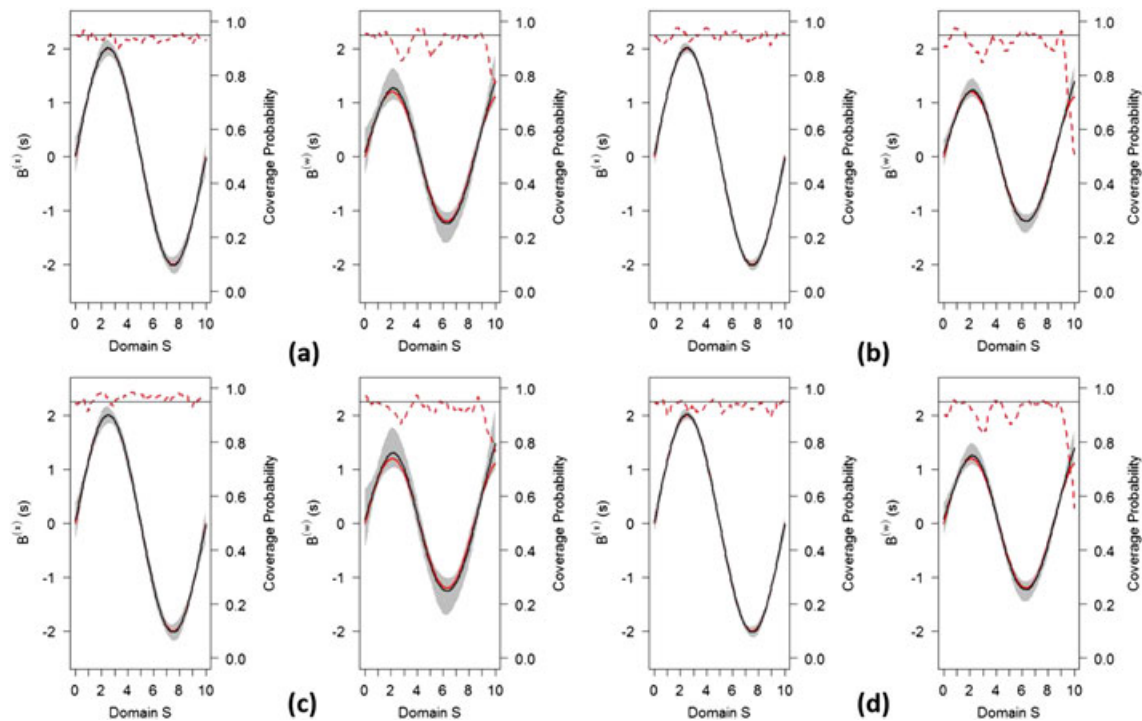


Figure 3. The functional joint model’s estimates of coefficient functions in Setting I. The true coefficient functions $B^{(x)}(s)$ (left plot in each panel) and $B^{(w)}(s)$ (right plot in each panel) and their estimates (black solid line), along with 95% confidence band (shaded region) and converge probability (red dotted line) with a reference line (horizontal line) at 0.95. Panel (a), $I = 200$, censoring rate (CR) = 0.3; panel (b), $I = 500$, CR = 0.3; panel (c), $I = 200$, CR = 0.5; panel (d), $I = 500$, CR = 0.5.

Section 4. The functional predictor for the i th subject $g_i(s)$ is simulated by $g_i(s) = \sum_{l=1}^{25} \xi_{il} \phi_l(s)$, for $i = 1, \dots, 400$, where ξ_{il} is the subject-specific principle component score generated from a normal distribution as $\xi_{il} \sim N(0, \lambda_l)$, where $l = 1, \dots, 25$. Thus, the simulated functional predictor $g_i(s)$ retains many features of the HRD. Next, we construct a 2D coefficient function (coefficient image) on the rectangle plane to match the size of the bean surface using the densities of bivariate normal distributions. Let

$$P_1 \sim N \left[\begin{pmatrix} 0.5 \\ 0.3 \end{pmatrix}, \begin{pmatrix} 0.025 & 0 \\ 0 & 0.015 \end{pmatrix} \right] \quad \text{and} \quad P_2 \sim N \left[\begin{pmatrix} 0.7 \\ 0.8 \end{pmatrix}, \begin{pmatrix} 0.03 & -0.01 \\ -0.01 & 0.03 \end{pmatrix} \right],$$

be density functions denoted by f_{P_1} and f_{P_2} , respectively. The coefficient image is given by $0.8f_{P_1} - 1.0f_{P_2}$, and its 1D representation $B(s)$ can be achieved by aligning points in the same order described previously. After both functional predictor and its coefficient are transformed to 1D domain, the longitudinal and survival data are generated via the same method as in Setting I. We select an approximate censoring rate of 30% and generate 200 simulated datasets. Our objective is to estimate the coefficient image via estimating its 1D representation $B(s)$ and mapping back to the 3D surfaces. Figure 4 displays the true and estimated coefficient functions mapped back to the bean surface. The close similarity of two figures indicates that the estimated coefficient function captures the main feature of the true coefficient function, suggesting the feasibility of our approach to transforming the high-dimensional surface image to the 1D domain.

6. Discussion

Functional data are increasingly collected in public health and medical studies to better understand important public health issues and complex diseases. Both theoretical and computational complexity in FDA often makes healthcare practitioners to reduce the rich functional data into several scalar measures, for example, volumes of a few brain regions. This enormous data reduction may distort the true underlying relationship between population’s health condition and the functional data. Moreover, some nontraditional functional data, such as genetic variant profiles defined along chromosomes or genomic regions, may not be able to reduce to scalar measures. To this end, FDA methods are increasingly used.

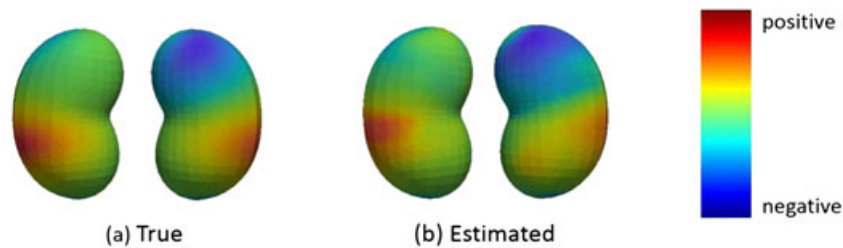


Figure 4. The true coefficient function (panel (a)) and estimated coefficient function (panel (b), the mean of 200 estimated $B(s)$) on a three-dimensional bean surface.

In this article, we develop an 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 the empirical orthonormal eigenfunctions obtained from FPCA. FPC scores can be readily included as scalar covariates in a standard joint model consisting of a linear mixed model for the longitudinal scalar response variable and a Cox model for the survival outcome. The parameters of the FJM are estimated in a maximum likelihood framework via EM algorithm. The proposed FJM provides a flexible framework to incorporate many features both in joint modeling of longitudinal and survival data and in FDA. We demonstrate through a simulation study that the FJM performs well in estimating the coefficient functions and other parameters.

Several studies have documented that diminishing hippocampal thickness at baseline is associated with an increased likelihood of progressing to clinical dementia [43]. However, these studies only assess changes to hippocampal volume rather than its surface morphology. Enormous information loss may result when aggregating the high-dimensional image data into a scalar volumetric value [44]. In comparison, our FJM framework accounts for the functional predictor (HRD) and other scalar covariates and efficiently estimate the association between the trajectory of cognitive functions measured by the ADAS-Cog 11 score and time to AD diagnosis. The inclusion of functional predictor HRD into the longitudinal submodel improves model fitting. We have identified that the regional radial atrophy in the CA1 subfield and the subiculum subfield is a good predictor of AD progression among patients with MCI. The identification of these subfields may facilitate case selection in clinical trials for evaluating therapeutic efficacy in slowing or modifying AD-related pathophysiology. Moreover, the proposed FJM can readily include multiple brain regions and even genotype profiles, as functional predictors to assess whether they are associated with AD progression.

There are some advantages of using FPC expansion, besides the convenience of FPC implementation. FPC expansion allows borrowing strength across subjects in estimating basis functions, and it can capture complex correlations within the functions, as in the ADNI study [45]. Moreover, FPC can easily handle observation with missing values in functions, or functions measured with errors [11]. We retain the first K eigenfunctions to approximate the functional predictor. Although the selected number of K eigenfunctions can explain the majority of the variability in the functional predictor, they may not adequately represent the coefficient function. Therefore, the features on the tail region of the coefficient function are not well captured as demonstrated in our simulation study. Furthermore, it is possible that some retained FPCs are not significantly associated with the outcome. On the other hand, in the functional predictor regression, there are various choices of basis functions, for example, splines, Fourier, wavelet, and their combinations, and regularization approaches. Splines are well suited to modeling simple and smooth functions and usually work better when the dimension of the grid is not too high [46]. However, no single basis function is superior in all settings. In the future work, we will investigate the performance of other basis functions and new methods for selecting among various candidate basis functions and regularization approaches.

There are some limitations that we will address in the future. First of all, we exclude from analysis the subjects with missing data in baseline covariates of interest. A majority of these subjects either do not have baseline image data measured or do not have image data in the archive because of the technical difficulty in image data collection and storage. We have assumed the baseline covariates are missing completely at random. We would like to investigate the validity of this assumption in the future. Moreover, there is some longitudinal MRI measurements available for some ADNI participants. About 30% of these participants have at least one missed MRI measurement, creating the missing data issue in longitudinal functional data. The complex within-subject correlation among functional data should be carefully modeled when

handling missing data. In addition, we use a time-invariant functional predictor in this article. It would be of scientific interest to extend the proposed FJM to accommodate longitudinal functional data. We can treat the longitudinal functional variable as a functional predictor, which can be decomposed by longitudinal FPCA [47] to account for its longitudinal data structure. Alternatively, the longitudinal functional variable can be treated 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 time to event of interest. Furthermore, in model (2), different formulations can be used to postulate how the hazard of a survival event depends on the longitudinal trajectory. For example, both the unobserved true value of $m_i(t_{ij})$ as in model (1) and its time-dependent slope $m'_i(t_{ij})$ can be included in model (2). A good summary of these various formulations in the joint modeling framework can be found in Rizopoulos *et al.* [48] and Yang *et al.* [49]. Finally, we would like to develop a user-friendly R package, which addresses the aforementioned issues and incorporates some useful features, for example, nonparametric smoothing and missing data.

Acknowledgements

Sheng Luo's research was supported in part by the National Institute of Neurological Disorders and Stroke under award numbers R01NS091307 and 5U01NS043127. 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 the 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.

References

1. Ramsay JO, Dalzell CJ. Some tools for functional data analysis. *Journal of the Royal Statistical Society. Series B (Methodological)* 1991; **53**(3):539–572.
2. Ramsay J, Silverman B. *Functional Data Analysis*. 1997. Springer: New York, 1997.
3. Marx BD, Eilers PH. Generalized linear regression on sampled signals and curves: a P-spline approach. *Technometrics* 1999; **41**(1):1–13.
4. Cardot H, Ferraty F, Sarda P. Functional linear model. *Statistics and Probability Letters* 1999; **45**(1):11–22.
5. Cardot H, Ferraty F, Sarda P. Spline estimators for the functional linear model. *Statistica Sinica* 2003; **13**(3):571–591.
6. James GM. Generalized linear models with functional predictors. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 2002; **64**(3):411–432.
7. Müller HG, Stadtmüller U. Generalized functional linear models. *The Annals of Statistics* 2005; **33**(2):774–805.
8. Reiss PT, Ogden RT. Functional principal component regression and functional partial least squares. *Journal of the American Statistical Association* 2007; **102**(479):984–996.
9. James GM, Wang J, Zhu J. Functional linear regression that's interpretable. *The Annals of Statistics* 2009; **37**(5A):2083–2108.
10. Goldsmith J, Bobb J, Crainiceanu CM, Caffo B, Reich D. Penalized functional regression. *Journal of Computational and Graphical Statistics* 2012; **20**(4):830–851.
11. Goldsmith J, Crainiceanu CM, Caffo B, Reich D. Longitudinal penalized functional regression for cognitive outcomes on neuronal tract measurements. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 2012; **61**(3):453–469.
12. Gertheiss J, Goldsmith J, Crainiceanu C, Greven S. Longitudinal scalar-on-functions regression with application to tractography data. *Biostatistics* 2013; **14**(3):447–461.
13. Gellar JE, Colantuoni E, Needham DM, Crainiceanu CM. Cox regression models with functional covariates for survival data. *Statistical Modelling* 2015; **15**(3):256–278.
14. Lee E, Zhu H, Kong D, Wang Y, Giovanello KS, Ibrahim JG. BFLCRM: a Bayesian functional linear Cox regression model for predicting time to conversion to Alzheimer's disease. *The Annals of Applied Statistics* 2015; **9**(4):2153–2178.
15. Faucett CL, Thomas DC. Simultaneously modelling censored survival data and repeatedly measured covariates: a Gibbs sampling approach. *Statistics in Medicine* 1996; **15**(15):1663–1685.
16. Wulfsohn MS, Tsiatis AA. A joint model for survival and longitudinal data measured with error. *Biometrics* 1997; **53**(1):330–339.
17. Tsiatis AA, Davidian M. Joint modeling of longitudinal and time-to-event data: an overview. *Statistica Sinica* 2004; **14**(3):809–834.
18. Proust-Lima C, Séne M, Taylor JMG, Jacqmin-Gadda H. Joint latent class models for longitudinal and time-to-event data: a review. *Statistical Methods in Medical Research* 2014; **23**(1):74–90.
19. Weiner MW, Veitch DP, Aisen PS, Beckett LA, Cairns NJ, Cedarbaum J, Green RC, Harvey D, Jack CR, Jagust W, Luthman J, Morris JC, Petersen RC, Saykin AJ, Shaw L, Shen L, Schwarz A, Toga AW, Trojanowski JQ, Alzheimer's Disease Neuroimaging Initiative. 2014 Update of the Alzheimer's Disease Neuroimaging Initiative: a review of papers published since its inception. *Alzheimer's and Dementia* 2015; **11**(6):e1–e120.
20. Li K, Chan W, Doody RS, Quinn J, Luo S. Prediction of conversion to Alzheimer's disease with longitudinal measures and time-to-event data. *Journal of Alzheimer's Disease* 2017; **58**(2):361–371.

21. Cleveland WS. Robust locally weighted regression and smoothing scatterplots. *Journal of the American Statistical Association* 1979; **74**(368):829–836.
22. Diggle P, Kenward MG. Informative drop-out in longitudinal data analysis. *Journal of the Royal Statistical Society. Series C (Applied Statistics)* 1994; **43**(1):49–93.
23. Lo RY, Hubbard AE, Shaw LM, Trojanowski JQ, Petersen RC, Aisen PS, Weiner MW, Jagust WJ, Alzheimer's Disease Neuroimaging Initiative. Longitudinal change of biomarkers in cognitive decline. *Archives of Neurology* 2011; **68**(10):1257–1266.
24. Zhang D, Shen D, Alzheimer's Disease Neuroimaging Initiative. Predicting future clinical changes of mci patients using longitudinal and multimodal biomarkers. *PLOS ONE* 2012; **7**(3):e33182.
25. Du AT. Magnetic resonance imaging of the entorhinal cortex and hippocampus in mild cognitive impairment and Alzheimer's disease. *Journal of Neurology, Neurosurgery and Psychiatry* 2001; **71**(4):441–447.
26. Apostolova LG, Mosconi L, Thompson PM, Green AE, Hwang KS, Ramirez A, Mistur R, Tsui WH, de Leon MJ. Sub-regional hippocampal atrophy predicts Alzheimer's dementia in the cognitively normal. *Neurobiology of Aging* 2010; **31**(7):1077–1088.
27. Qiu A, Fennema-Notestine C, Dale AM, Miller MI, Alzheimer's Disease Neuroimaging Initiative. Regional shape abnormalities in mild cognitive impairment and Alzheimer's disease. *NeuroImage* 2009; **45**(3):656–661.
28. Thompson PM, Hayashi KM, De Zubicaray GI, Janke AL, Rose SE, Semple J, Hong MS, Herman DH, Gravano D, Doddrell DM, Toga AW. Mapping hippocampal and ventricular change in Alzheimer disease. *NeuroImage* 2004; **22**(4):1754–1766.
29. Yushkevich PA. Continuous medial representation of brain structures using the biharmonic PDE. *NeuroImage* 2009; **45**(1):S99–S110.
30. Shi J, Thompson PM, Gutman B, Wang Y, Initiative ADN, et al. Surface fluid registration of conformal representation: application to detect disease burden and genetic influence on hippocampus. *NeuroImage* 2013; **78**:111–134.
31. Staniswalis JG, Lee JJ. Nonparametric regression analysis of longitudinal data. *Journal of the American Statistical Association* 1998; **93**(444):1403–1418.
32. Yao F, Müller H-G, Clifford AJ, Dueker SR, Follett J, Lin Y, Buchholz BA, Vogel JS. Shrinkage estimation for functional principal component scores with application to the population kinetics of plasma folate. *Biometrics* 2003; **59**(3):676–685.
33. Crainiceanu CM, Staicu A-M, Ray S, Punjabi N. Bootstrap-based inference on the difference in the means of two correlated functional processes. *Statistics in Medicine* 2012; **31**(26):3223–3240.
34. Crainiceanu CM, Reiss PT, Goldsmith J, Huang L, Huo L, Scheipl F. Refund: regression with functional data 2013.
35. Peng J, Paul D. A geometric approach to maximum likelihood estimation of the functional principal components from sparse longitudinal data. *Journal of Computational and Graphical Statistics* 2009; **18**(4):995–1015.
36. Rizopoulos D. JM: An R package for the joint modelling of longitudinal and time-to-event data. *Journal of Statistical Software* 2010; **35**(9):1–33.
37. Risacher SL, Saykin AJ, Wes JD, Shen L, Firpi HA, McDonald BC. Baseline MRI predictors of conversion from MCI to probable ad in the ADNI cohort. *Current Alzheimer Research* 2009; **6**(4):347–361.
38. Cui Y, Liu B, Luo S, Zhen X, Fan M, Liu T, Zhu W, Park M, Jiang T, Jin JS, Alzheimer's Disease Neuroimaging Initiative. Identification of conversion from mild cognitive impairment to Alzheimer's disease using multivariate predictors. *PLOS ONE* 2011; **6**(7):e21896.
39. Li S, Okonkwo O, Albert M, Wang Mei-Cheng. Variation in variables that predict progression from MCI to AD dementia over duration of follow-up. *American Journal of Alzheimer's Disease* 2013; **2**(1):12–28.
40. Corder EH, Saunders AM, Strittmatter WJ, Schmechel DE, Gaskell PC, Small GWet, Roses A.D, Haines JL, Pericak-Vance MA. Gene dose of apolipoprotein E type 4 allele and the risk of Alzheimer's disease in late onset families. *Science* 1993; **261**(5123):921–923.
41. Apostolova LG, Dutton RA, Dinov ID, Hayashi KM, Toga AW, Cummings JL, Thompson PM. Conversion of mild cognitive impairment to Alzheimer disease predicted by hippocampal atrophy maps. *Archives of Neurology* 2006; **63**(5):693–699.
42. Frankó E, Joly O, Alzheimer's Disease Neuroimaging Initiative. Evaluating Alzheimer's disease progression using rate of regional hippocampal atrophy. *PLOS ONE* 2013; **8**(8):e71354.
43. Weiner MW, Veitch DP, Aisen PS, Beckett LA, Cairns NJ, Green RC, Harvey D, Jack CR, Jagust W, Liu E, Morris JC, Petersen RC, Saykin AJ, Schmidt ME, Shaw L, Shen L, Siuclak JA, Soares H, Toga AW, Trojanowski JQ, Alzheimer's Disease Neuroimaging Initiative. The Alzheimer's disease neuroimaging initiative: a review of papers published since its inception. *Alzheimer's & Dementia* 2013; **9**(5):e111–e194.
44. Wang Y, Song Y, Rajagopalan P, An T, Liu K, Chou Y-Y, et al. Surface-based TBM boosts power to detect disease effects on the brain: an N=804 ADNI study. *NeuroImage* 2011; **56**(4):1993–2010.
45. Yao F, Müller H-G, Wang J-L. Functional data analysis for sparse longitudinal data. *Journal of the American Statistical Association* 2005; **100**(470):577–590.
46. Morris JS. Functional regression. *Annual Review of Statistics and Its Application* 2015; **2**(1):321–359.
47. Greven S, Crainiceanu C, Caffo B, Reich D. Longitudinal functional principal component analysis. *Electronic Journal of Statistics* 2010; **4**:1022–1054.
48. Rizopoulos D, Hatfield LA, Carlin BP, Takkenberg JJM. Combining dynamic predictions from joint models for longitudinal and time-to-event data using Bayesian model averaging. *Journal of the American Statistical Association* 2014; **109**(508):1385–1397.
49. Yang L, Yu M, Gao S. Prediction of coronary artery disease risk based on multiple longitudinal biomarkers. *Statistics in Medicine* 2016; **35**(8):1299–1314.

Supporting information

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