Check for updates

Received: 11 June 2020 | Revised: 6 November 2020 | Accepted: 8 January 2021

DOI: 10.1111/biom.13427

BIOMETRIC METHODOLOGY

Biometrics A JOURNAL OF THE INTERNATIONAL BIOMETRIC SOCIETY WILEY

Joint model for survival and multivariate sparse functional data with application to a study of Alzheimer's Disease

Cai Li[1] | **Luo Xiao**[2] | **Sheng Luo**[3]

[1] Department of Biostatistics, Yale University, New Haven, Connecticut, USA

[2] Department of Statistics, North Carolina State University, Raleigh, North Carolina, USA

[3] Department of Biostatistics and Bioinformatics, Duke University, Durham, North Carolina, USA

Correspondence Cai Li, Department of Biostatistics, Yale University, New Haven, CT 06511, USA.
Email: cai.li@yale.edu

Funding information
NIH, Grant/Award Numbers: R56 AG064803, R01 AG064803, R01 NS112303

Abstract

Studies of Alzheimer's disease (AD) often collect multiple longitudinal clinical outcomes, which are correlated and predictive of AD progression. It is of great scientific interest to investigate the association between the outcomes and time to AD onset. We model the multiple longitudinal outcomes as multivariate sparse functional data and propose a functional joint model linking multivariate functional data to event time data. In particular, we propose a multivariate functional mixed model to identify the shared progression pattern and outcome-specific progression patterns of the outcomes, which enables more interpretable modeling of associations between outcomes and AD onset. The proposed method is applied to the Alzheimer's Disease Neuroimaging Initiative study (ADNI) and the functional joint model sheds new light on inference of five longitudinal outcomes and their associations with AD onset. Simulation studies also confirm the validity of the proposed model. Data used in preparation of this article were obtained from the ADNI database.

KEYWORDS
EM algorithm, functional mixed model, multivariate longitudinal data, smoothing, survival

1 | INTRODUCTION

Alzheimer's disease (AD) is the most prevalent neurodegenerative disorder, can often be characterized by accelerated metal degradation over time, and may ultimately progress to dementia. In the year of 2017, AD was the sixth leading cause of death in the United States with 121,494 recorded deaths (Alzheimer's Association, 2019). Great efforts have been dedicated to advancing early detection of AD.

The motivating data are from the Alzheimer's Disease Neuroimaging Initiative (ADNI) with the primary goal of investigating whether multimodal data can be combined to measure the progression of AD (Weiner *et al.*, 2017). We are interested in jointly modeling clinical variables, multiple longitudinal outcomes measured intermittently and time to AD onset or drop-out. Throughout the paper, AD onset refers to a clinical declaration of probable AD based on cognitive symptoms. We consider five longitudinal biomarkers commonly measured in AD studies. Among the five biomarkers, high values of Disease Assessment Scale-Cognitive 13 items (ADAS-Cog 13) and Functional Assessment Questionnaire (FAQ) reflect severe cognitive decline, whereas low values of Rey Auditory Verbal Learning Test immediate recall (RAVLT-immediate), Rey Auditory Verbal Learning Test learning curve (RAVLT-learn) and Mini-Mental State Examination (MMSE) indicate a high risk for developing AD.

Figure 1 presents spaghetti plots of the five longitudinal biomarkers and highlights profiles and time to AD onset for two subjects. Subject A has a more acute cognitive decline compared to Subject B. Indeed, subject A has faster increasing values of ADAS-Cog 13 and FAQ, faster decreasing values of RAVLT-immediate, RAVLT-learn, and MMSE

 © 2021 The International Biometric Society

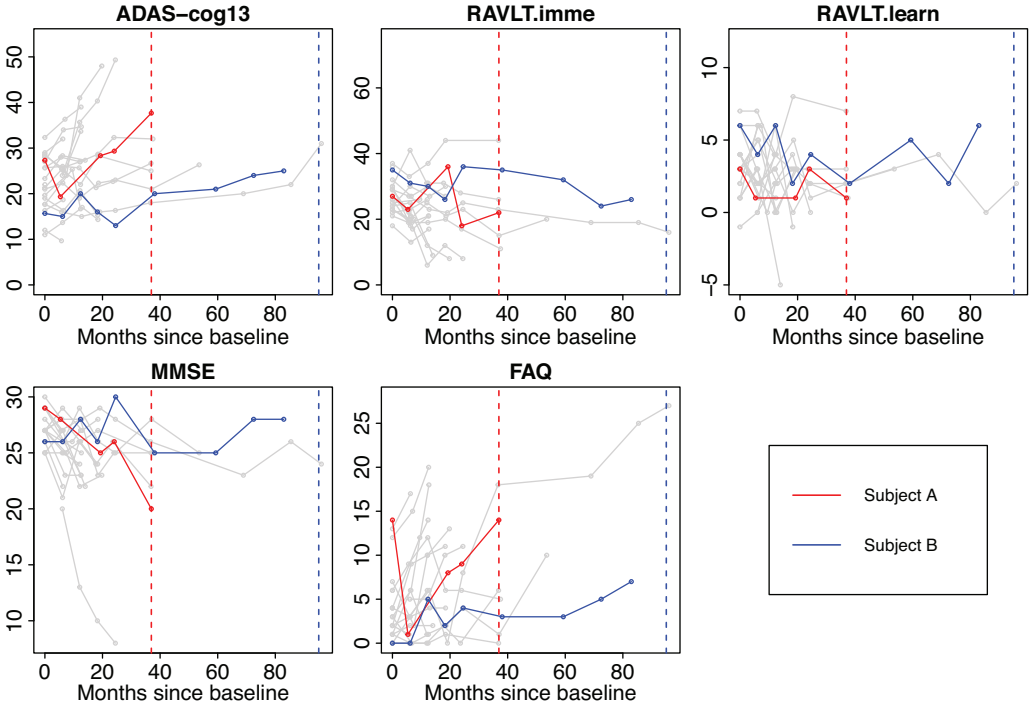

FIGURE 1 Five observed longitudinal biomarkers of two subjects. Vertical lines: time to AD onset; gray lines: longitudinal outcomes for 20 other subjects

as well as an earlier AD onset around month 40. The plots motivate our research: (a) subjects with more severe cognitive impairment seem more likely to progress to AD; and (b) the multiple longitudinal outcomes may be correlated. In addition, the observations do not always show clear linear trends and the trends across outcomes may not be synchronized, suggesting of potentially heterogeneous patterns.

Parametric models are often used in the joint modeling literature (Tsiatis and Davidian, 2004). De Gruttola and Tu (1994) introduced a shared random effects model, which links longitudinal data to event time data via a set of time-invariant random effects. Wulfsohn and Tsiatis (1997) proposed a shared latent process model, which models the instantaneous effect of longitudinal data to event time data. Extensions of the univariate joint models to multiple longitudinal outcomes include Henderson *et al.* (2000) and Lin *et al.* (2002). However, limitations exist: (1) the parametric models are incapable of modeling complex nonlinear trends of longitudinal outcomes; (2) assuming a specific structure of the correlation may be subject to model misspecification. Recently, joint modeling of event time data and functional data has drawn some attention. Yao (2007) proposed a shared latent process model. Yan *et al.* (2017) proposed a shared random effect model with a two-step estimation, and Ye *et al.* (2015) proposed a model for baseline longitudinal patterns and interval-censored event time data. To clarify, joint modeling here refers to the situation where the domain of function is longitudinal time. By contrast, Cox regression models where functional data are used as baseline covariates have been extensively studied; see Kong *et al.* (2018) and references therein. Nevertheless, none of the above works considered joint modeling of multivariate functional data and event time data.

To capture the heterogeneity of patterns in the outcomes as well as correlations among them, a popular method is multivariate functional principal component analysis (Happ and Greven, 2018). Then a Cox regression model may be adopted to link the outcomes and AD onset via the functional principal component scores. However, there exist serious computational issues for joint model estimation. Multivariate functional principal component analysis (MFPCA) models the mean functions, auto-covariance functions for within-function correlations, and cross-covariances for between-function correlations nonparametrically. Thus, for J-dimensional functional data, there are J univariate functions and J^2 bivariate functions to estimate. If we use tensor product splines to approximate the bivariate functions and each marginal basis is of dimension c, then it leads to J^2c^2 parameters to estimate, which is computationally prohibitive for joint model estimation if J is more than 2 and infeasible for large J. Moreover, MFPCA is mostly used for dimension reduction and the resulting multivariate eigenfunctions and associated scores are often difficult to interpret. In

particular, MFPCA does not explicitly model the correlation between outcomes as is often done in parametric models for multivariate longitudinal data (Verbeke *et al.*, 2014).

We propose a new multivariate functional mixed model (MFMM) for multivariate functional data and subsequently a new functional joint model for linking multivariate functional data and event time data. The advantages of the proposed methods include: (1) the MFMM retains the flexibility of functional data methods for capturing nonlinear patterns in longitudinal outcomes and models the correlation between outcomes via a shared latent process; (2) compared to J^2 bivariate covariance functions for MFPCA, MFMM requires only two bivariate covariance functions, which makes joint model estimation feasible; (3) MFMM explicitly separates the shared latent process, common to all outcomes, from the outcome-specific latent processes, unique to each outcome, and thus greatly enhances model interpretability; (4) MFMM enables a flexible Cox regression model, which not only evaluates the effects of the shared latent process to survival risk but also identifies the additional contribution of each outcome to survival risk.

The remainder of the paper is organized as follows. Section 2 introduces the proposed MFMM as well as the joint model for disease progression and survival. Section 3 describes a two-step estimation method and the proposed joint estimation method. Section 4 presents model selection for the proposed model. Section 5 applies the proposed model to the ADNI data. Section 6 examines the numerical properties of the proposed method through simulations. Section 7 concludes this work with some discussion. Technical details and extra results for numerical studies are included in the Web Appendices.

2 | MODEL

Let Y_{ijk} be the kth observation of the jth outcome (biomarker) measured intermittently at time t_{ijk} for subject i with $1 \le k \le m_{ij}$, $1 \le j \le J$, and $1 \le i \le n$. So there are n subjects and subject i has m_{ij} observations for the jth outcome. In our data application, we model the five longitudinal outcomes ADAS-Cog 13, RAVLT-immediate, RAVLT-learn, MMSE, and FAQ (plotted in Figure 1) as Y_{ijk} with $j = 1, \dots, 5$, respectively. The survival time for the ith subject is denoted by S_i and is assumed to be subject to independent right censoring with censoring time denoted by C_i. Let $T_i = \min(S_i, C_i)$ and $\Delta_i = 1_{\{S_i \le C_i\}}$. Note that $t_{ijk} \in [0, T_i]$, meaning no observations after T_i. Assume $T_i \le \tau$ for all i, where τ is the length of study follow-up. Denote by $\mathbf{z}_i = (Z_{i1}, \dots, Z_{iP}) \in \mathbb{R}^P$ the vector of baseline covariates.

2.1 | Multivariate functional mixed model

We propose a MFMM for the multivariate latent process $\{X_{i1}(t), \dots, X_{iJ}(t)\}$. The MFMM is of the form

$$Y_{ijk} = X_{ij}(t_{ijk}) + \epsilon_{ijk}, \quad X_{ij}(t) = \mu_j(t) + \beta_j\{U_i(t) + W_{ij}(t)\}, \tag{1}$$

where ϵ_{ijk} are random noises so that the longitudinal outcome Y_{ijk} is a proxy observation of the true latent stochastic process $X_{ij}(t)$ evaluated at time t_{ijk}. The smooth latent process $X_{ij}(t)$ is decomposed into three components. First, $\mu_j(t)$ is the fixed mean function for outcome j. For simplicity, we assume that the mean function only depends on the longitudinal time but it may depend on the baseline covariates, which can be incorporated easily using, for example, additive models. The continuous latent profile $U_i(t)$, common to multiple outcomes, is a subject-specific random deviation from the mean functions. $U_i(t)$ captures the subject-specific disease progression pattern and correlation among outcomes. It represents subject i's unique latent disease status at time t manifested by multiple outcomes and can be specified so that a higher value indicates more severe status. The outcome-specific scaling parameter β_j is the expected increase in outcome j for one unit increase in $U_i(t)$. If two outcomes are negatively correlated, their β_js have different signs. $W_{ij}(t)$ is the subject- and outcome-specific random deviation from the outcome-specific mean, and it characterizes subject i's outcome-specific progression pattern. By multiplying the scaling parameters β_j, $W_{ij}(t)$ are comparable across outcomes.

We model $U_i(\cdot)$ and $W_{ij}(\cdot)$ via two zero-mean Gaussian processes with covariance functions $C_0(s,t) = \mathrm{Cov}\{U_i(s), U_i(t)\}$ and $C_1(s,t) = \mathrm{Cov}\{W_{ij}(s), W_{ij}(t)\}$, respectively. Consider the spectral decomposition of the covariance functions, $C_0(s,t) = \sum_\ell d_{0\ell}\phi_\ell(s)\phi_\ell(t)$ and $C_1(s,t) = \sum_\ell d_{1\ell}\psi_\ell(s)\psi_\ell(t)$, where $d_{01} \ge d_{02} \ge \cdots$ and $d_{11} \ge d_{12} \ge \cdots$ are the ordered eigenvalues, and $\phi_\ell(\cdot)$ and $\psi_\ell(\cdot)$ are the associated orthonormal eigenfunctions satisfying $\int_0^\tau \phi_\ell(t)\phi_{\ell'}(t)dt = \int_0^\tau \psi_\ell(t)\psi_{\ell'}(t)dt = 1_{\{\ell=\ell'\}}$. Then the Karhunen–Loève representations of $U_i(t)$ and $W_{ij}(t)$ are $U_i(t) = \sum_{\ell\ge1}\phi_\ell(t)\xi_{i\ell}$, $W_{ij}(t) = \sum_{\ell\ge1}\psi_\ell(t)\zeta_{ij\ell}$, where $\xi_{i\ell} \sim \mathcal{N}(0, d_{0\ell})$ are eigen scores and independent over ℓ, and $\zeta_{ij\ell} \sim \mathcal{N}(0, d_{1\ell})$ are defined similarly and independent over j and ℓ. The eigenfunctions $\phi_\ell(t)$ and $\psi_\ell(t)$ represent the changing patterns of the latent disease profiles, and the random scores $\xi_{i\ell}$ and $\zeta_{ij\ell}$ determine how strongly subject i's latent disease profile follows those patterns. In practice, we assume there are only a finite number of patterns so that

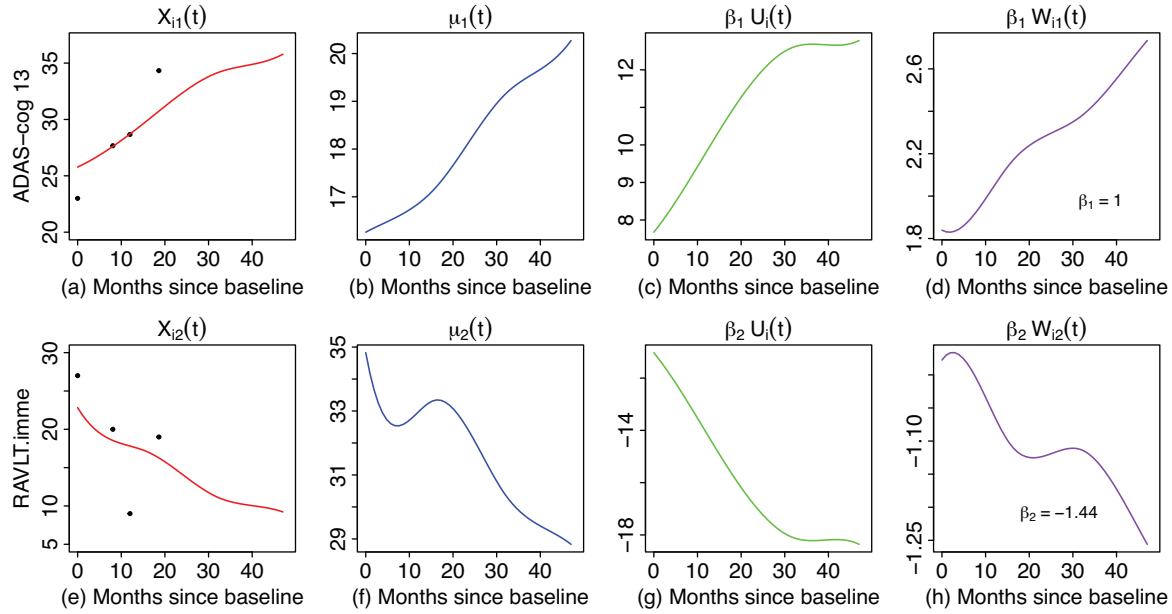

FIGURE 2 Estimate of each component in model (1) for outcomes ADAS-Cog 13 (A–D) and RAVLT-immediate (E–H) for one subject in the ADNI study. (A) and (E): observed ADAS-Cog 13 and RAVLT-immediate values (black dots) and the latent processes $X_{ij}(t)$; (B) and (F): mean functions $\mu_j(t)$; (C) and (G): shared latent disease profile $U_i(t)$ multiplied by β_j ; (D) and (H): outcome-specific deviations $W_{ij}(t)$ multiplied by β_j

$U_i(t) = \sum_{\ell=1}^{L_0} \phi_\ell(t)\xi_{i\ell}, W_{ij}(t) = \sum_{\ell=1}^{L_1} \psi_\ell(t)\zeta_{ij\ell}$, where L_0 and L_1 are finite numbers. We shall treat L_0 and L_1 as tuning parameters and select them through data adaptive methods; see Section 4 for details. We assume that the random noises ϵ_{ijk} are independent and normally distributed with zero mean and variance σ_j^2. Finally, $U_i(t)$, $W_{ij}(t)$, and ϵ_{ijk} are assumed independent between subjects and across each other.

We illustrate the proposed MFMM by fitting five biomarkers in the ADNI study and present two outcomes ADAS-Cog 13 and RAVLT-immediate in Figure 2; see Section 5 for model fitting details. In Figure 2, the estimate or prediction of each component in model (1) is visualized for the two outcomes of one subject. To make model identifiable, we set $\beta_1 = 1$ for the outcome ADAS-Cog 13. Figures 2(A) and 2(E) present the fitted latent processes by MFMM for both outcomes. The subject shows steady worsening in cognitive function $X_{i1}(t)$ (increasing ADAD-Cog 13 in Figure 2(A)), which can be decomposed into increasing mean cognitive function $\mu_1(t)$ (Figure 2(B)), deteriorating (increasing) latent disease profile $U_i(t)$ (Figure 2(C)), and positive outcome-specific progression $W_{i1}(t)$ (Figure 2(D)). Similar interpretation can be made to the outcome RAVLT-immediate and decreasing patterns indicate AD progression.

Model (1) allows us to explicitly model the shared latent disease profile $U_i(t)$ between the outcomes and outcome-specific profile $W_{ij}(t)$. While it has a similar multilevel decomposition structure as in multilevel FPCA (Di *et al.*, 2009), there exist significant differences. The proposed MFMM accommodates outcome heterogeneity: (1) The scaling parameters reduce heterogeneity of the multiple functions, which may measure quantitatively very different features of subjects. For example, β_2 for RAVLT-immediate is estimated as a negative number and it changes the direction of $U_i(t)$ so that it is negatively correlated with ADAS-Cog 13; and (2) the outcome-specific progression further accommodates data heterogeneity, such as, $\beta_1 W_{i1}(t)$ shows larger deviation from zero toward AD onset as compared with $\beta_2 W_{i2}(t)$, suggesting more severe disease progression in ADAS-Cog 13 than in RAVLT-immediate of the subject. Compared with multivariate FPCA, MFMM borrows its idea but further accounts for outcome-specific patterns, which leads to theoretical and practical advantages: (1) MFMM gives a more interpretable model of multiple outcomes by separating the shared component, which models the correlation between outcomes, from outcome-specific components, which model the patterns of outcomes that are uncorrelated from other outcomes. By contrast, MFPCA only considers the shared component by reducing the multiple outcomes into a set of uncorrelated scores; and (2) by imposing a parsimonious model, the number of auto- and cross-covariance functions to be estimated is not increasing with the number of outcomes. This is a reasonable compromise between computability and medical fidelity,

which greatly alleviates computational burden and thus makes the joint estimation feasible. In addition, MFMM may be regarded as a nonparametric extension of parametric multilevel decomposition model for multivariate longitudinal data (Verbeke *et al.*, 2014).

We derive from (1) that

$$C_{jj'}(s,t) := \mathrm{Cov}\{X_{ij}(s), X_{ij'}(t)\} = \beta_j\beta_{j'}C_0(s,t) + \beta_j^2 1_{\{j=j'\}}C_1(s,t). \quad (2)$$

If $j \neq j'$, $C_{jj'}(s,t) = \beta_j\beta_{j'}C_0(s,t)$. For model identifiability, we let $\beta_1 = 1$. Then it can be shown that for $J \geq 2$, β_j and $C_0(\cdot,\cdot)$ can be uniquely determined by (2) using the equations with $j \neq j'$. (See Web Appendix A for proofs and Web Appendix B for the covariance structure relating MFMM to MFPCA.)

2.2 | Joint model for disease progression and survival

To model the survival time, we use the proportional hazards model

$$h_i(t) = h_0(t)\exp\{\mathbf{z}_i^\top\boldsymbol{\gamma}_z + \mathcal{F}(\mathbf{x}_i,t)\}, \quad (3)$$

where $h_0(\cdot)$ is the baseline hazard function, $\boldsymbol{\gamma}_z$ is the coefficient vector corresponding to baseline covariates $\mathbf{z}_i$, $\mathbf{x}_i$ is the collection of all outcomes for subject i, and $\mathcal{F}(\mathbf{x}_i,t)$ is the regression term of multiple latent processes at time t. We consider the framework of shared random effects models (Wu and Carroll, 1988; De Gruttola and Tu, 1994) as it takes into account the entire history of the latent processes and let $\mathcal{F}(\mathbf{x}_i,t) = \sum_{\ell=1}^{L_0}\xi_{i\ell}\gamma_{0\ell} + \sum_{j=1}^{J}\sum_{\ell=1}^{L_1}\zeta_{ij\ell}\gamma_{1j\ell}$, where $\gamma_{0\ell}$ and $\gamma_{1j\ell}$ are the coefficients corresponding to the shared and outcome-specific latent profiles, respectively. The hazard model extends the model in Yan *et al.* (2017) for univariate functional data to multivariate functional data.

2.3 | Likelihood of joint model

For model estimation, we now derive the likelihood function of the multivariate longitudinal outcomes and the event time data.

We shall introduce some notation, which will be used throughout the rest of the paper. Let $\boldsymbol{\xi}_i = (\xi_{i1},\ldots,\xi_{iL_0})^\top$ be the vector of scores for the shared latent profile $U_i(t)$. Then $\boldsymbol{\xi}_i \sim \mathcal{N}(\mathbf{0},\mathbf{D}_0)$, where $\mathbf{D}_0 \in \mathbb{R}^{L_0\times L_0}$ is a diagonal matrix with $d_{0\ell}$ the ℓth diagonal element. Let $\boldsymbol{\zeta}_{ij} = (\zeta_{ij1},\ldots,\zeta_{ijL_1})^\top$ be the vector of scores for the outcome-specific latent profiles $W_{ij}(t)$. Then $\boldsymbol{\zeta}_{ij} \sim \mathcal{N}(\mathbf{0},\mathbf{D}_1)$, where $\mathbf{D}_1 \in \mathbb{R}^{L_1\times L_1}$ is a diagonal matrix with $d_{1\ell}$ the ℓth diagonal element. Similarly, let $\mathbf{y}_{ij} = (Y_{ij1},\ldots,Y_{ijm_i})^\top$ be the vector of observations for the jth outcome and $\mathbf{y}_i = (\mathbf{y}_{i1}^\top,\ldots,\mathbf{y}_{iJ}^\top)^\top$. Let $\boldsymbol{\mu}_{ij} = \{\mu_{ij}(t_{i1}),\ldots,\mu_{ij}(t_{im_i})\}^\top$ be the vector of the jth mean function at the observed time points. Let $\boldsymbol{\Phi}(t) = \{\phi_1(t),\ldots,\phi_{L_0}(t)\}^\top$ and $\boldsymbol{\Psi}(t) = \{\psi_1(t),\ldots,\psi_{L_1}(t)\}^\top$. Then let $\boldsymbol{\Phi}_i = \{\boldsymbol{\Phi}(t_{i1}),\ldots,\boldsymbol{\Phi}(t_{im_i})\}^\top$ and $\boldsymbol{\Psi}_i = \{\boldsymbol{\Psi}(t_{i1}),\ldots,\boldsymbol{\Psi}(t_{im_i})\}^\top$ be the matrices of eigenfunctions evaluated at the observed time points. Denote by $\mathbf{x}_{ij} = \{X_{ij}(t_{i1}),\ldots,X_{ij}(t_{im_i})\}^\top$ the vector of the jth outcome evaluated at the observed time points without measurement errors, note that $\mathbf{x}_{ij} = \beta_j(\boldsymbol{\Phi}_i\boldsymbol{\xi}_i + \boldsymbol{\Psi}_i\boldsymbol{\zeta}_{ij})$, and $\mathbf{x}_i = (\mathbf{x}_{i1}^\top,\ldots,\mathbf{x}_{iJ}^\top)^\top$. Finally, let $\mathbf{t}_i = (t_{i1},\ldots,t_{im_i})^\top$ be the vector of the observed time points, and $\boldsymbol{\Sigma}_i = \mathrm{blockdiag}(\sigma_1^2\mathbf{I}_{m_i},\ldots,\sigma_J^2\mathbf{I}_{m_i})$.

First, the conditional likelihood of multivariate longitudinal data is

$$f(\mathbf{y}_i|\mathbf{x}_i,\mathbf{t}_i,\boldsymbol{\Sigma}_i) = (|2\pi\boldsymbol{\Sigma}_i|)^{-\frac{1}{2}}\exp\left\{-\frac{1}{2}(\mathbf{y}_i-\mathbf{x}_i)^\top\boldsymbol{\Sigma}_i^{-1}(\mathbf{y}_i-\mathbf{x}_i)\right\} \quad (4)$$

and $f(\boldsymbol{\xi}_i|\mathbf{D}_0) = (|2\pi\mathbf{D}_0|)^{-\frac{1}{2}}\exp(-\frac{1}{2}\boldsymbol{\xi}_i^\top\mathbf{D}_0^{-1}\boldsymbol{\xi}_i)$, $f(\boldsymbol{\zeta}_{ij}|\mathbf{D}_1) = (|2\pi\mathbf{D}_1|)^{-\frac{1}{2}}\exp(-\frac{1}{2}\boldsymbol{\zeta}_{ij}^\top\mathbf{D}_1^{-1}\boldsymbol{\zeta}_{ij})$. Next, the conditional likelihood of time-to-event data is given by

$$f(T_i,\Delta_i|h_0,\mathbf{z}_i,\mathbf{x}_i,\boldsymbol{\gamma}_z,\boldsymbol{\gamma}_\eta) = \left\{h_0(T_i)\exp(\mathbf{z}_i^\top\boldsymbol{\gamma}_z + \boldsymbol{\eta}_i^\top\boldsymbol{\gamma}_\eta)\right\}^{\Delta_i} \exp\left\{-\int_0^{T_i}h_0(u)\exp(\mathbf{z}_i^\top\boldsymbol{\gamma}_z + \boldsymbol{\eta}_i^\top\boldsymbol{\gamma}_\eta)du\right\}, \quad (5)$$

where $\boldsymbol{\eta}_i = (\boldsymbol{\xi}_i^\top,\boldsymbol{\zeta}_{i1}^\top,\ldots,\boldsymbol{\zeta}_{iJ}^\top)^\top$, $\boldsymbol{\gamma}_\eta = (\boldsymbol{\gamma}_0^\top,\boldsymbol{\gamma}_{11}^\top,\ldots,\boldsymbol{\gamma}_{1J}^\top)^\top$, $\boldsymbol{\gamma}_0 = (\gamma_{01},\ldots,\gamma_{0L_0})^\top$, and $\boldsymbol{\gamma}_{1j} = (\gamma_{1j1},\ldots,\gamma_{1jL_1})^\top$ for all j. As the multivariate longitudinal data and the time-to-event data are conditionally independent given the latent process $\mathbf{x}_i$, the marginal likelihood is given by

$$\prod_{i=1}^{n}\left[\int f(\mathbf{y}_i|\mathbf{x}_i,\mathbf{t}_i,\boldsymbol{\Sigma}_i)f(\boldsymbol{\xi}_i|\mathbf{D}_0)\left\{\prod_{j=1}^{J}f(\boldsymbol{\zeta}_{ij}|\mathbf{D}_1)\right\} \times f(T_i,\Delta_i|h_0,\mathbf{z}_i,\mathbf{x}_i,\boldsymbol{\gamma}_z,\boldsymbol{\gamma}_\eta)d\boldsymbol{\eta}_i\right]. \quad (6)$$

3 | MODEL ESTIMATION

3.1 | Two-Step method

A naive estimation method would be to use a two-step method by first predicting the scores from the longitudinal

data model (1) and then doing plug-in for the Cox regression model (3). The first step is nontrivial, so we shall provide some details.

First, for each longitudinal biomarker, the mean function is estimated by penalized splines (Eilers and Marx, 1996). Next, we adopt the fast covariance estimation method for multivariate sparse functional data in Li *et al.* (2020) to obtain estimates of the auto- and cross-covariance functions, $\widehat{C}_{jj'}$, and error variances, $\widehat{\sigma}_j^2$, via bivariate penalized splines. Finally, we treat the estimates as true auto- and cross-covariances and estimate β_j as in Web Appendix A. Once $\widehat{\beta}_j$ are obtained, $\widehat{C}_0$ can be solved by least squares using Equation (2) with $j \neq j'$. Then, $\widehat{C}_1$ can be solved similarly using the same equation with $j = j'$. The negative eigenvalues will be discarded to ensure that the covariances are positive semidefinite. We then use the conditional expectation approach for predicting the scores, a popular approach in traditional joint modeling (Wulfsohn and Tsiatis, 1997) and sparse functional data analysis (Yao *et al.*, 2005); see Web Appendix C for details. Finally, the predicted scores, $\widehat{\mathbb{E}}(\boldsymbol{\eta}_i|\mathbf{y}_i)$, where the estimates of fixed quantities are plugged in, will be used in the Cox regression.

Despite its computational advantage, the two-step method has some well-known drawbacks: (1) it is a marginal approach that ignores the inherent correlation between the longitudinal and survival process and often leads to inferior statistical efficiency; (2) the predicted scores in the first step are usually biased, and the estimation error will propagate into the subsequent Cox regression. Nevertheless, we shall compare the two-step method with the proposed estimation method below and demonstrate the superiority of the latter one in the numerical study. In addition, the estimates from the two-step method can be used as initial values for the joint estimation method.

3.2 | Monte Carlo EM method

3.2.1 | Reduced rank splines

Following Yao (2007) and Huang *et al.* (2014), we use reduced rank splines for modeling the smooth mean functions and covariance functions. Let $\mathbf{b}(t) = \{B_1(\cdot), \ldots, B_c(\cdot)\}^\top$ be the vector of B-spline basis functions in the unit interval (de Boor, 1978), where c is the number of equally spaced interior knots plus the order (degree plus 1) of the B-splines. We model the mean function $\mu_j(t)$ by $\mathbf{b}(t)^\top \boldsymbol{\alpha}_j$, where $\boldsymbol{\alpha}_j$ is the coefficient vector of the jth mean function. Let $\mathbf{G} = \int \mathbf{b}(t)\mathbf{b}(t)^\top dt \in \mathbb{R}^{c\times c}$, which is positive definite (Zhou *et al.*, 1998). Then $\widetilde{\mathbf{b}}(t) = \mathbf{G}^{-\frac{1}{2}}\mathbf{b}(t)$ are orthonormal B-spline bases. For the covariance functions C_0 and C_1, we approximate their ℓth eigenfunctions $\phi_\ell(t)$ and $\psi_\ell(t)$ by $\widetilde{\mathbf{b}}(t)^\top \boldsymbol{\theta}_{0\ell}$ and $\widetilde{\mathbf{b}}(t)^\top \boldsymbol{\theta}_{1\ell}$, respectively, where $\boldsymbol{\theta}_{0\ell}$ and $\boldsymbol{\theta}_{1\ell}$ are coefficient vectors. Let $\boldsymbol{\Theta}_0 = [\boldsymbol{\theta}_{01}, \ldots, \boldsymbol{\theta}_{0L_0}]$ and $\boldsymbol{\Theta}_1 = [\boldsymbol{\theta}_{11}, \ldots, \boldsymbol{\theta}_{1L_1}]$. Then the orthonormality of eigenfunctions gives the constraints, $\boldsymbol{\Theta}_0^\top \boldsymbol{\Theta}_0 = \mathbf{I}_{L_0\times L_0}$ and $\boldsymbol{\Theta}_1^\top \boldsymbol{\Theta}_1 = \mathbf{I}_{L_1\times L_1}$. These constraints are equivalent to $\boldsymbol{\theta}_{0\ell}^\top \boldsymbol{\theta}_{0\ell'} = \boldsymbol{\theta}_{1\ell}^\top \boldsymbol{\theta}_{1\ell'} = 1_{\{\ell=\ell'\}}$.

3.2.2 | E-Step

Although nonparametric functions are components of the proposed model, their spline representations allow a parametric estimation based on the EM algorithm. The full data likelihood depends on the latent random variables $\boldsymbol{\eta}_i$ and can be optimized via the EM method, which treats $\boldsymbol{\eta}_i$ as missing values and iterates between E-steps and M-steps until convergence. Such a strategy is often deployed in parametric joint modeling (Wulfsohn and Tsiatis, 1997). We shall use the Monte Carlo EM algorithm, an alternative to the Gaussian–Hermite quadrature, to approximate the numerical integrals in the E-step.

Let $\boldsymbol{\alpha} = (\boldsymbol{\alpha}_1^\top, \ldots, \boldsymbol{\alpha}_J^\top)^\top$ be the vector of spline coefficients for the mean functions, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_J)^\top$ the vector of scaling parameters and $\boldsymbol{\sigma}^2 = (\sigma_1^2, \ldots, \sigma_J^2)^\top$ the vector of error variances. Denote by $\boldsymbol{\Omega} = \{h_0, \boldsymbol{\beta}, \boldsymbol{\gamma}_z, \boldsymbol{\gamma}_\eta, \mathbf{D}_0, \mathbf{D}_1, \boldsymbol{\alpha}, \boldsymbol{\Theta}_0, \boldsymbol{\Theta}_1, \boldsymbol{\sigma}^2\}$ the set of parameters and $\widehat{\boldsymbol{\Omega}} = \{\widehat{h}_0, \widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\gamma}}_z, \widehat{\boldsymbol{\gamma}}_\eta, \widehat{\mathbf{D}}_0, \widehat{\mathbf{D}}_1, \widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{\Theta}}_0, \widehat{\boldsymbol{\Theta}}_1, \widehat{\boldsymbol{\sigma}}^2\}$ the estimate. Let $g(\cdot)$ be any smooth function of $\boldsymbol{\eta}_i$, then the conditional expectation $\mathbb{E}\{g(\boldsymbol{\eta}_i)|T_i, \Delta_i, \mathbf{z}_i, \mathbf{y}_i, \mathbf{t}_i, \widehat{\boldsymbol{\Omega}}\}$ is given by

$$\frac{\int g(\boldsymbol{\eta}_i) f(T_i, \Delta_i|\widehat{h}_0, \mathbf{z}_i, \boldsymbol{\eta}_i, \widehat{\boldsymbol{\gamma}}_z, \widehat{\boldsymbol{\gamma}}_\eta) f(\boldsymbol{\eta}_i|\mathbf{y}_i, \mathbf{t}_i, \widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\Theta}}_0, \widehat{\boldsymbol{\Theta}}_1, \widehat{\mathbf{D}}_0, \widehat{\mathbf{D}}_1, \widehat{\boldsymbol{\sigma}}^2) d\boldsymbol{\eta}_i}{\int f(T_i, \Delta_i|\widehat{h}_0, \mathbf{z}_i, \boldsymbol{\eta}_i, \widehat{\boldsymbol{\gamma}}_z, \widehat{\boldsymbol{\gamma}}_\eta) f(\boldsymbol{\eta}_i|\mathbf{y}_i, \mathbf{t}_i, \widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\Theta}}_0, \widehat{\boldsymbol{\Theta}}_1, \widehat{\mathbf{D}}_0, \widehat{\mathbf{D}}_1, \widehat{\boldsymbol{\sigma}}^2) d\boldsymbol{\eta}_i},$$

where $f(T_i, \Delta_i|\widehat{h}_0, \mathbf{z}_i, \boldsymbol{\eta}_i, \widehat{\boldsymbol{\gamma}}_z, \widehat{\boldsymbol{\gamma}}_\eta)$ is the conditional likelihood in (5), and the second part of the denominator can be obtained from the joint normality of $\boldsymbol{\eta}_i$ and $\mathbf{y}_i$, given the data and parameter estimates; see Web Appendix C. We now use $\mathbb{E}_i\{g(\boldsymbol{\eta}_i)\}$ to denote the conditional expectation for convenience. In the E-step, because the integrals for the conditional expectations have no closed form solution, we use Monte Carlo approximation

$$\mathbb{E}_i\{g(\boldsymbol{\eta}_i)\} \approx \frac{\sum_{q=1}^{Q} g(\boldsymbol{\eta}_i^{(q)}) f(T_i, \Delta_i|\widehat{h}_0, \mathbf{z}_i, \boldsymbol{\eta}_i^{(q)}, \widehat{\boldsymbol{\gamma}}_z, \widehat{\boldsymbol{\gamma}}_\eta)}{\sum_{q=1}^{Q} f(T_i, \Delta_i|\widehat{h}_0, \mathbf{z}_i, \boldsymbol{\eta}_i^{(q)}, \widehat{\boldsymbol{\gamma}}_z, \widehat{\boldsymbol{\gamma}}_\eta)},$$

where $\boldsymbol{\eta}_i^{(q)}$ is the qth sample from the normal distribution $f(\boldsymbol{\eta}_i|\mathbf{y}_i, \mathbf{t}_i, \widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\Theta}}_0, \widehat{\boldsymbol{\Theta}}_1, \widehat{\mathbf{D}}_0, \widehat{\mathbf{D}}_1, \widehat{\boldsymbol{\sigma}}^2)$, and Q random samples are drawn. To accelerate the convergence, we use the

estimates from the two-step method as the initial values of the parameters.

3.2.3 | M-Step

Estimates of the current iteration can be obtained by optimizing separate parts of the joint likelihood (6) in the M-step, because each part only involves disjoint sets of parameters. Specifically, $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, $\boldsymbol{\Theta}_0$, $\boldsymbol{\Theta}_1$, and $\boldsymbol{\sigma}^2$ can be estimated iteratively by minimizing the expected negative log likelihood of the longitudinal process (4),

$$\sum_{i=1}^{n}\sum_{j=1}^{J}\left[\frac{m_i}{2}\log(2\pi\sigma_j^2)+\frac{1}{2\sigma_j^2}\mathbb{E}_i\left\{\mathbf{y}_{ij}-\mathbf{B}_i\boldsymbol{\alpha}_j-\beta_j\left(\widetilde{\mathbf{B}}_i\boldsymbol{\Theta}_0\boldsymbol{\xi}_i+\widetilde{\mathbf{B}}_i\boldsymbol{\Theta}_1\boldsymbol{\zeta}_{ij}\right)\right\}^2\right].$$

We adopt an iterative algorithm to cyclically estimate the columns of $\boldsymbol{\Theta}_0$ and $\boldsymbol{\Theta}_1$ and deploy an ad hoc step to satisfy the orthonormality constraints on the parameter matrices. The parameters in the diagonal matrices $\mathbf{D}_0$ and $\mathbf{D}_1$ are estimated by minimizing the expected negative logarithm of $\prod_{i=1}^{n} f(\boldsymbol{\xi}_i|\mathbf{D}_0)$ and $\prod_{i=1}^{n}\prod_{j=1}^{J} f(\boldsymbol{\zeta}_{ij}|\mathbf{D}_1)$, respectively. The baseline hazard function h_0 and the parameter vectors $\boldsymbol{\gamma}_z$ and $\boldsymbol{\gamma}_\eta$ in the Cox regression can be estimated according to the expected negative log likelihood of the survival process (5),

$$\sum_{i=1}^{n}\left[-\Delta_i\left\{\log h_0(T_i)+\mathbf{z}_i^\top\boldsymbol{\gamma}_z+\mathbb{E}_i(\boldsymbol{\eta}_i^\top\boldsymbol{\gamma}_\eta)\right\}+\int_0^{T_i}h_0(u)\times\mathbb{E}_i\left\{\exp\left(\mathbf{z}_i^\top\boldsymbol{\gamma}_z+\boldsymbol{\eta}_i^\top\boldsymbol{\gamma}_\eta\right)\right\}du\right].$$

In particular, the baseline hazard h_0 is estimated nonparametrically by the Breslow estimator, and $\boldsymbol{\gamma}_z$ and $\boldsymbol{\gamma}_\eta$ are updated by a one-step Newton–Raphson algorithm inside the loop. The estimated standard errors of the Cox regression coefficients can be obtained by inverting the observed information matrix. We defer the technical details to Web Appendix D.

4 | MODEL SELECTION

As described in Section 3.2, cubic B-splines are used for approximating mean functions and eigenfunctions. We use equally spaced knots for constructing the splines and for simplicity, we use the same number of knots for all spline functions. Following Huang *et al.* (2014), we use the asymptotic theory in Li and Hsing (2010) to determine the number of basis functions according to the sample size. For the simulation and data application, we use nine spline bases ($c = 9$) which is found to work well; see Web Appendix E for implementation details.

The number of eigenfunctions is an important tuning parameter since it determines the functional characteristics of the latent stochastic process. We use information criteria for model selection, which requires an evaluation of the model complexity, the degrees of freedom of the model. It can be shown that the negative log likelihood is given by $\ell_n = -2\sum_{i=1}^{n}[\log f(\mathbf{y}_i|\widehat{\boldsymbol{\Omega}})+\log\mathbb{E}\{f(T_i,\Delta_i|\mathbf{z}_i,\boldsymbol{\eta}_i,\widehat{\boldsymbol{\Omega}})\}]$, where $f(\mathbf{y}_i|\widehat{\boldsymbol{\Omega}})$ is a normal density with the covariance described in Web Appendix C, and the expectation can be approximated by $Q^{-1}\sum_{q=1}^{Q} f(T_i,\Delta_i|\widehat{h}_0,\mathbf{z}_i,\boldsymbol{\eta}_i^{(q)},\widehat{\boldsymbol{\gamma}}_z,\widehat{\boldsymbol{\gamma}}_\eta)$.

We approximate the degrees of freedom via the number of effective parameters,

$$\mathrm{df} := Jc+(L_0+L_1)(c+1)+P+L_0+JL_1+2J-1-\frac{L_0(L_0+1)}{2}-\frac{L_1(L_1+1)}{2},$$

where Jc is the number of parameters for estimating the mean functions, $(L_0+L_1)(c+1)$ is corresponding to the eigen pairs, $P+L_0+JL_1$ is the number of coefficients in Cox regression, $2J-1$ is corresponding to the error variance σ_j^2 and the scaling factor β_j, and the last two terms are due to orthonormality constraints on $\boldsymbol{\Theta}_0$ and $\boldsymbol{\Theta}_1$, the matrix of spline coefficients for eigenfunctions (see Section 3.2.1). Therefore, we may calculate AIC $= \ell_n + 2\cdot\mathrm{df}$, and BIC $= \ell_n + \log n\cdot\mathrm{df}$. We shall use a two-dimensional grid for selecting the two tuning parameters L_0 and L_1.

5 | DATA ANALYSIS

We apply the proposed functional joint model (denoted as FJM) to the ADNI data for jointly characterizing the varying patterns of the multivariate longitudinal outcomes and their association with time to diagnosis of AD. The data are from the first two phases of ADNI, which contain 803 participants with amnestic mild cognitive impairment (MCI, a transition risk state between normal state and AD state) at baseline who had at least one follow-up visit. Participants of the first phase were scheduled to be assessed at baseline, 6, 12, 18, 24, and 36 months with additional annual follow-ups included in the second phase. Note that the exact follow-up times can actually vary. Thus, for the combined data, the average number of visits is 4.72. For the analysis, the following variables are used as baseline covariates: baseline age (mean: 74.4, standard deviation: 7.3, range 55.1–89.3), gender (36.1% female), years of education (mean: 15.6, standard deviation: 3.0, range 4–20), and the number of apolipoprotein E $\epsilon 4$ alleles (APOE4, 56% $\geqq$ 1), given their potential effects on AD progression (Fleisher *et al.*, 2007).

TABLE 1 Model comparison. The "best model" row gives selected number(s) of eigenfunctions by BIC

	FJM	Reduced A	Reduced B	Reduced C	MJM
$X_{ij}(t)$	MFMM	MFMM	MFMM	$\mu_j(t)+\beta_j U_i(t)$	$\beta_{0j}+\beta_{1j}t+b_{0ij}+b_{1ij}t$
$\mathcal{F}(\mathbf{x}_i,t)$	$\sum_{\ell=1}^{L_0}\xi_{i\ell}\gamma_{0\ell}+\sum_{j=1}^{J}\sum_{\ell=1}^{L_1}\zeta_{ij\ell}\gamma_{1j\ell}$	$\sum_{\ell=1}^{L_0}\xi_{i\ell}\gamma_{0\ell}$	$\sum_{j=1}^{J}\sum_{\ell=1}^{L_1}\zeta_{ij\ell}\gamma_{1j\ell}$	$\sum_{\ell=1}^{L_0}\xi_{i\ell}\gamma_{0\ell}$	$\sum_{j=1}^{J}(b_{0ij}+b_{1ij}t)\gamma_j$
Best model	$(L_0,L_1)=(2,2)$	$(L_0,L_1)=(2,2)$	$(L_0,L_1)=(2,2)$	$L_0=2$	NaN
log likelihood	−37773.30	−37849.41	−37954.25	−39718.01	−47249.77
AIC	75754.60	75906.83	76116.50	79590.03	94657.54
BIC	76242.19	76394.42	76604.09	79951.03	95272.71
Concordance	0.86	0.85	0.77	0.84	0.85

We consider various models listed in Table 1, including the proposed functional joint model and its variants. The hazard model is specified as

$$h_i(t)=h_0(t)\exp\left\{\text{Age}_i\gamma_a+\text{Gender}_i\gamma_g+\text{Education}_i\gamma_e+\text{APOE4}_i\gamma_\epsilon+\mathcal{F}(\mathbf{x}_i,t)\right\}$$

and the form of $\mathcal{F}(\mathbf{x}_i,t)$ is given in Table 1. Reduced models A and B share the same MFMM submodel (1) as FJM. Reduced model A is a special case of the proposed model with only shared components $\boldsymbol{\xi}_i$ contributing to survival risk. Reduced model B is another special case with only outcome-specific components $\boldsymbol{\zeta}_{ij}$ contributing to survival risk. Compared to FJM, reduced model C only has shared components $\boldsymbol{\xi}_i$ in both submodels. To fit FJM, we use the settings of splines described in Section 4. The number of Monte Carlo samples and stopping criteria are set as described in Web Appendix E. We select the numbers of eigenfunctions according to BIC as we shall show in Section 6 that it performs well for model selection. The three reduced models can be estimated similarly as the full FJM using the proposed MCEM approach, and model selection can be similarly carried out using BIC. In addition, we consider the parametric multivariate joint linear model (denoted as MJM) proposed by Henderson *et al.* (2000) and implemented in `R` package `joineRML` (Hickey *et al.*, 2018). With slight abuse of notation, we denote by β_{0j} and β_{1j} the fixed effects, and b_{0ij} and b_{1ij} the random intercept and slope in a linear mixed effects model for the jth outcome. For MJM, the corresponding Cox coefficient for the jth outcome is γ_j.

Table 1 presents the overall performance of the various models. First, FJM has the highest likelihood and smallest AIC and BIC, which compares favorably against the other models. Second, Reduced model A is the closest to FJM in terms of the three criteria, followed by Reduced model B and Reduced model C. We shall see later that the shared components play a major role in determining AD risk, and hence Reduced model A outperforms Reduced model B. Reduced model C not only overlooks the outcome-specific components in Cox regression, but also ignores that heterogeneity in modeling longitudinal outcomes, which explains why its performance is inferior to Reduced models A and B. Furthermore, MJM is outperformed by all other models, indicating that it gains to model the longitudinal outcomes nonlinearly. Finally, we include the concordance index (Harrell, 2015) for evaluating the predictive ability of survival models as an additional criterion. Again, FJM has the highest concordance index, while other models are slightly inferior to it. It is not surprising to see that Reduced model B is ranked last among the competitors since the shared components are primary contributors to hazard risk.

As suggested by one reviewer, a partial functional linear model (PFL) might be adopted proposed if the survival part is the primary interest (Kong *et al.*, 2016). PFL treats the functional outcomes as cross-sectional covariates and uses a linear combination of eigen scores of each functional variable as the predictor in Cox regression. Despite the potential multicollinearity of the scores, this model is similar to using $W_{ij}(t)$ in MFMM but does not separate the shared component from the outcome-specific components. One primary objective of joint modeling is to understand the associations between features of the longitudinal outcomes and time to disease progression (Tsiatis and Davidian, 2004), and it is known that multicollinearity may be an issue for this objective. Nonetheless, we have compared MFMM, MFPCA, and PFL for survival prediction and found that MFMM performs best in terms of the concordance index and PFL is outperformed by MFPCA. The above models are also compared for fitting the longitudinal outcomes further showing advantages of MFMM; see Web Appendix F for details.

Table 2 summarizes the estimated Cox coefficients from the functional joint model. We have the following remarks. (1) The results show that APOE4 is significantly associated with AD risk at level 0.05, which is consistent with existing AD studies. In particular, the presence of APOE4 allele increases the hazard of AD diagnosis by 39.10% while

TABLE 2 Estimates (standard errors) of Cox regression coefficients from functional joint model. An asterisks indicates significance at level 0.05

FJM	Coefficient	Estimate (standard error)	*p*-value
Age	γ_a	−0.01 (0.01)	0.06
Gender (female)	γ_g	0.27 (0.23)	0.25
Education	γ_e	0.03 (0.03)	0.29
APOE4	γ_ϵ	0.33 (0.15)*	0.03
Shared latent progression	γ_{01}	0.33 (0.02)*	$3e-51$
	γ_{02}	0.29 (0.09)*	0.01
ADAS-Cog 13 progression	γ_{11}	0.00 (0.08)	0.97
	γ_{12}	−0.24 (0.22)	0.28
RAVLT-immediate progression	γ_{21}	0.20 (0.08)*	0.01
	γ_{22}	0.73 (0.19)*	$9e-5$
RAVLT-learn progression	γ_{31}	0.03 (0.08)	0.74
	γ_{32}	−0.06 (0.29)	0.83
MMSE progression	γ_{41}	−0.04 (0.06)	0.51
	γ_{42}	−0.26 (0.28)	0.35
FAQ progression	γ_{51}	0.09 (0.07)	0.16
	γ_{52}	−0.21 (0.20)	0.29

adjusting for other covariates. (2) The parameters γ_{01}, γ_{02} capture the effects of the latent disease process $U_i(t)$ manifested by the five biomarkers. The significance of these effects indicates the contribution of the latent profile shared among the longitudinal outcomes to the hazard of AD conversion, after adjusting for baseline clinical covariates. The result agrees with the excellent predictive performance of the five biomarker reported in Li *et al.* (2017). (3) The proposed functional joint model sheds new insight on the AD study by successfully identifying important associations between individual longitudinal outcomes and the survival. In Table 2, the individual effects of RAVLT-immediate γ_{21}, γ_{22} are significant, while others are not. These results suggest that the progression patterns of the longitudinal outcomes (ADAS-Cog 13, RAVLT-learn, MMSE, and FAQ) contribute to AD diagnosis mainly through the shared latent profile, not through their outcome-specific progression. By contrast, RAVLT-immediate contributes through the outcome-specific progression in addition to the shared latent progression. Our findings are again supported by an independent study (Li *et al.*, 2019), which applied a penalized method and consistently selected RAVLT-immediate as the only significant risk factor of AD conversion.

Figure 3 presents the estimated mean functions of longitudinal outcomes by three methods. Both FJM and MJM are based on joint estimation and show a similar trend: the mean curves are progressing toward mental deterioration over the months, suggesting an increased risk of developing AD. These findings confirm the intuition since the participants in the study suffer from MCI, which causes cognitive decline toward dementia. Moreover, FJM can further characterize the nonlinear pattern of the biomarkers. Although MJM only provides linear estimates, it correctly identifies the deteriorating trend. The two-step method (denoted as 2-step) fails in capturing such a degenerate trend since the curves are relatively stable during the study period; this might be because the estimates are biased due to its marginal nature.

Additional results for the ADNI data are presented in Web Appendix F.

6 | SIMULATIONS

6.1 | Simulation settings

In this section, we compare the performance of the joint estimation, and the two-step method, of the proposed model. We consider two cases of data generation and replicate each for 100 times. Here, we focus on case 1, which is a realistic setting with data generated according to the fitted model of the ADNI study in Web Appendix G; see Web Appendix H for case 2 which is an alternative setting, but its results show a similar pattern as case 1.

The longitudinal data are generated according to MFMM (1) with two outcomes $J = 2$. The outcome-specific

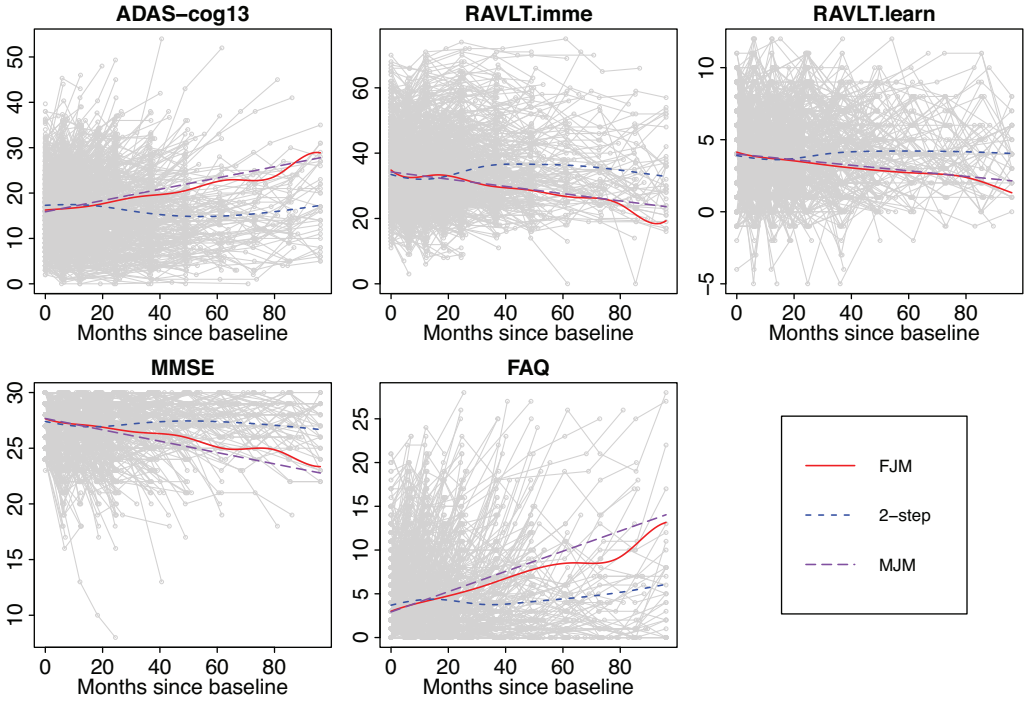

FIGURE 3 Estimated mean functions. Gray lines: longitudinal outcomes

mean functions, scaling parameters, and error variances are derived from the estimates of ADAS-Cog 13 and RAVLT-immediate. We set two principal components for both two covariances, and the eigenfunctions are specified as the estimates of $C_0(s,t)$ and $C_1(s,t)$. The eigen scores $\xi_{i\ell}$s are generated from a normal distribution $\mathcal{N}(0, d_{0\ell})$ with $d_{01} = 95.41$ and $d_{02} = 5.04$. The outcome-specific eigen scores $\zeta_{ij\ell}$s are generated similarly with $d_{11} = 21.90$ and $d_{12} = 2.05$. We set the scaling parameters $\beta_1 = 1$ and $\beta_2 = -1.44$. The white noise ϵ_{ijk}s are sampled from a normal distribution $\mathcal{N}(0, \sigma_j^2)$, where $\sigma_1^2 = 9.49$ and $\sigma_2^2 = 21.98$. The observed time points $t_{ijk} = t_{ik}$ are 11 fixed time points of the ADNI study mapped to the interval [0,1].

The time-to-event data are generated according to Cox regression (3) with the coefficients set as the estimates of the common components and outcome-specific components of ADAS-Cog 13 and RAVLT-immediate. We use the baseline hazard function $h_0(t) = 1$ and specify the linear hazard rate function as $\sum_{\ell=1}^{2} \xi_{i\ell}\gamma_{0\ell} + \sum_{j=1}^{2}\sum_{\ell=1}^{2} \zeta_{ij\ell}\gamma_{1j\ell}$, where the Cox coefficients are $\boldsymbol{\gamma}_0 = (0.33, 0.31)^\top$, $\boldsymbol{\gamma}_{11} = (0.01, -0.27)^\top$, and $\boldsymbol{\gamma}_{12} = (0.25, 0.80)^\top$. Then failure times are drawn independently from a standard exponential distribution. Censoring times C_is are generated independently from a uniform distribution on $[0, c_0]$, where c_0 is a constant and the final truncation time $\tau = 1$ is used so that the censoring rate is around 65%. For each subject, only measurements at $t_{ik} \le T_i$ are retained. We generate data with 803 subjects, and the average number of observations per subject is around 5.5. All of these settings are close to the real case of the ADNI study.

6.2 | Simulation results

We use the settings of model fitting described in Web Appendix E. First, we fix the number of principal components as the truth $L_0 = L_1 = 2$ and estimate model components. In most of the replications, FJM converges within 200 iterations. The first two rows of Figure 4 present the estimated mean functions and eigenfunctions for $\phi_2(t)$ and $\psi_2(t)$. While the medians of FJM are close to the truth, the two-step method has significant bias over the time. Furthermore, we obtain the point-wise confidence bands based on the quantiles of all the replications. The 95% confidence bands of FJM are able to cover the truth, but this is not the case for the two-step method since the true means lie outside its 95% confidence bands. The last two rows of Figure 4 summarize the estimates of the Cox coefficients. FJM is reasonably close to the truth, but the two-step method shows significant bias.

Finally, we use AIC and BIC to select the number of eigenfunctions in the covariances and evaluate the performance of the proposed approaches. One may define the candidate ranks of the covariance by the proportion of variance explained in the marginal MFMM stage since the two-step method is used for providing initial values. For the two-step method, the number of principal components

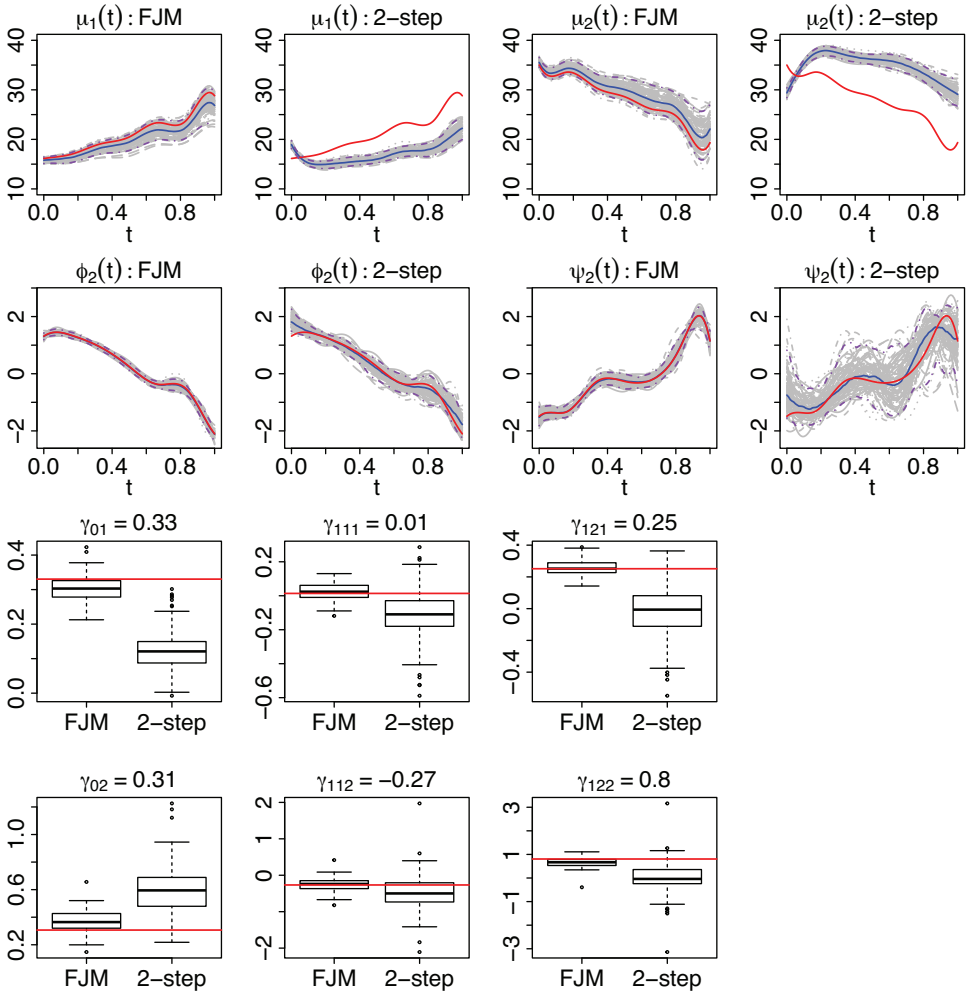

FIGURE 4 Estimated functions/parameters of 100 replications. The first row: estimated mean functions; the second row: estimated eigenfunctions; the last two rows: estimated Cox coefficients. Red lines: true functions/parameters; gray lines: estimated functions; blue lines: medians of estimates; dashed purple lines: 95% point-wise confidence bands

can be selected by using either AIC or BIC solely based on Cox regression as in Kong *et al.* (2018). For the two-step method, the rates of correctly selecting two principal components for the covariances are 0.40 and 0.38 using BIC, respectively. By contrast, the proposed approach for FJM achieves excellence in practice, the correct selection rates are 1.00 for all using BIC. For both the two methods, the rates of AIC are slightly lower than those of BIC, so we use BIC for rank selection in the data application.

In summary, FJM shows very competitive performance and is superior to the two-step method in terms of estimation and rank selection in all scenarios. Additional simulation results are included in Web Appendix H.

7 | DISCUSSION

Our work can be extended in several directions. First, the MFMM framework is flexible to further account for heterogeneity across multiple longitudinal outcomes. For example, one might use two different scaling parameters multiplying $U_i(t)$ and $W_{ij}(t)$ in model (1). Moreover, one might model $W_{ij}(t)$ with heterogeneous covariances to incorporate any prior information. We have adopted model (1) with homogeneous covariances in this paper as it is found to better fit the ADNI data. Second, it is worth developing a joint integrative modeling framework to incorporate imaging, genetic, and longitudinal biomarkers into the Cox regression (3) and comparing it with a predictive model proposed by Kong *et al.* (2015), which treats time to AD as the survival outcomes and uses multimodal data to predict AD progression. Finally, the theoretic properties of the proposed joint model are unclear, which warrant avenues for future research.

ACKNOWLEDGEMENTS

The authors thank the editor, the associate editor, and the three reviewers for their constructive and helpful

comments, which greatly improved the article. This work was partially supported by the NIH grants R56 AG064803 (CL, LX, and SL), R01 AG064803 (LX and SL), and R01 NS112303 (LX). Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (http://adni.loni.usc.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 report. A complete listing of ADNI investigators can be found at http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf.

DATA AVAILABILITY STATEMENT

The Alzheimer's Disease Neuroimaging Initiative data that support the findings in this paper are openly available at http://adni.loni.usc.edu.

ORCID

Cai Li https://orcid.org/0000-0002-5624-3031
Luo Xiao https://orcid.org/0000-0001-8707-0914
Sheng Luo https://orcid.org/0000-0003-4214-5809

REFERENCES

Alzheimer's Association, (2019) 2019 Alzheimer's disease facts and figures. *Alzheimer's & Dementia*, 15, 321–387.

de Boor, C. (1978) *A Practical Guide to Splines*. Berlin, Germany: Springer.

De Gruttola, V. and Tu, X.M. (1994) Modelling progression of CD4-lymphocyte count and its relationship to survival time. *Biometrics*, 50, 1003–1014.

Di, C.-Z., Crainiceanu, C.M., Caffo, B.S. and Punjabi, N.M. (2009) Multilevel functional principal component analysis. *The Annals of Applied Statistics*, 3, 458–488.

Eilers, P. and Marx, B. (1996) Flexible smoothing with B-splines and penalties (with discussion). *Statistical Science*, 11, 89–121.

Fleisher, A., Sowell, B., Taylor, C., Gamst, A., Petersen, R.C., Thal, L. et al. (2007) Clinical predictors of progression to Alzheimer disease in amnestic mild cognitive impairment. *Neurology*, 68, 1588–1595.

Happ, C. and Greven, S. (2018) Multivariate functional principal component analysis for data observed on different (dimensional) domains. *Journal of the American Statistical Association*, 113, 649–659.

Harrell Jr, F.E. (2015) *Regression Modeling Strategies: with Applications to Linear Models, Logistic and Ordinal Regression, and Survival Analysis*. Berlin, Germany: Springer.

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

Hickey, G.L., Philipson, P., Jorgensen, A. and Kolamunnage-Dona, R. (2018) joineRML: a joint model and software package for time-to-event and multivariate longitudinal outcomes. *BMC Medical Research Methodology*, 18, 50.

Huang, H., Li, Y., and Guan, Y. (2014) Joint modeling and clustering paired generalized longitudinal trajectories with application to cocaine abuse treatment data. *Journal of the American Statistical Association*, 109, 1412–1424.

Kong, D., Giovanello, K. S., Wang, Y., Lin, W., Lee, E., Fan, Y., et al. (2015) Predicting Alzheimer's disease using combined imaging-whole genome SNP data. *Journal of Alzheimer's Disease*, 46, 695–702.

Kong, D., Ibrahim, J.G., Lee, E. and Zhu, H. (2018) FLCRM: functional linear Cox regression model. *Biometrics*, 74, 109–117.

Kong, D., Xue, K., Yao, F. and Zhang, H.H. (2016) Partially functional linear regression in high dimensions. *Biometrika*, 103, 147–159.

Li, C., Xiao, L. and Luo, S. (2020) Fast covariance estimation for multivariate sparse functional data. *Stat*, 9, e245.

Li, K., Chan, W., Doody, R.S., Quinn, J. and Luo, S. (2017) Prediction of conversion to Alzheimer's disease with longitudinal measures and time-to-event data. *Journal of Alzheimer's Disease*, 58, 361–371.

Li, S., Wu, Q. and Sun, J. (2019) Penalized estimation of semiparametric transformation models with interval-censored data and application to Alzheimer's disease. *Statistical Methods in Medical Research*, 29, 2151–2166.

Li, Y. and Hsing, T. (2010) Uniform convergence rates for nonparametric regression and principal component analysis in functional/longitudinal data. *The Annals of Statistics*, 38, 3321–3351.

Lin, H., McCulloch, C.E. and Mayne, S.T. (2002) Maximum likelihood estimation in the joint analysis of time-to-event and multiple longitudinal variables. *Statistics in Medicine*, 21, 2369–2382.

Tsiatis, A.A. and Davidian, M. (2004) Joint modeling of longitudinal and time-to-event data: An overview. *Statistica Sinica*, 14, 809–834.

Verbeke, G., Fieuws, S., Molenberghs, G. and Davidian, M. (2014) The analysis of multivariate longitudinal data: A review. *Statistical Methods in Medical Research*, 23, 42–59.

Weiner, M.W., Veitch, D.P., Aisen, P.S., Beckett, L.A., Cairns, N.J., Green, R.C. et al. (2017) Recent publications from the Alzheimer's disease neuroimaging initiative: reviewing progress toward improved ad clinical trials. *Alzheimer's & Dementia*, 13, e1–e85.

Wu, M.C. and Carroll, R.J. (1988) Estimation and comparison of changes in the presence of informative right censoring by modeling the censoring process. *Biometrics*, 44, 175–188.

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

Yan, F., Lin, X. and Huang, X. (2017) Dynamic prediction of disease progression for leukemia patients by functional principal component analysis of longitudinal expression levels of an oncogene. *The Annals of Applied Statistics*, 11, 1649–1670.

Yao, F. (2007) Functional principal component analysis for longitudinal and survival data. *Statistica Sinica*, 17, 965–983.

Yao, F., Müller, H.-G. and Wang, J.-L. (2005) Functional data analysis for sparse longitudinal data. *Journal of the American Statistical Association*, 100, 577–590.

Ye, J., Li, Y. and Guan, Y. (2015) Joint modeling of longitudinal drug using pattern and time to first relapse in cocaine dependence treatment data. *The Annals of Applied Statistics*, 9, 1621–1642.

Zhou, S., Shen, X. and Wolfe, D. (1998) Local asymptotics for regression splines and confidence regions. *The Annals of Statistics*, 26, 1760–1782.

SUPPORTING INFORMATION

Web Appendices A, B, C, D, E, F, G, and H referenced in Section 2, 3, 4, 5, and 6 and code implementing the proposed method are available with this paper at the Biometrics website on Wiley Online Library.

How to cite this article: Li C, Xiao L, Luo S. Joint model for survival and multivariate sparse functional data with application to a study of Alzheimer's Disease. *Biometrics*. 2021;1–13. https://doi.org/10.1111/biom.13427