RESEARCH ARTICLE

Statistics in Medicine WILEY

Bayesian inference and dynamic prediction of multivariate joint model with functional data: An application to Alzheimer's disease

Haotian Zou[1] | Kan Li[2] | Donglin Zeng[1] | Sheng Luo[3] | Alzheimer's Disease Neuroimaging Initiative[1]

[1]Department of Biostatistics, Gillings School of Global Public Health, CB#7420, Duke University, Chapel Hill, North Carolina, USA

[2]Merck Research Lab, Merck & Co, North Wales, Pennsylvania, USA

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

Correspondence
Sheng Luo, Department of Biostatistics and Informatics, Duke University, 2424 Erwin Rd, Durham, NC 27705, USA
Email: sheng.luo@duke.edu

Funding information
National Institute on Aging, Grant/Award Number: R01AG064803

Alzheimer's disease (AD) is a severe neurodegenerative disorder impairing multiple domains, for example, cognition and behavior. Assessing the risk of AD progression and initiating timely interventions at early stages are critical to improve the quality of life for AD patients. Due to the heterogeneous nature and complex mechanisms of AD, one single longitudinal outcome is insufficient to assess AD severity and disease progression. Therefore, AD studies collect multiple longitudinal outcomes, including cognitive and behavioral measurements, as well as structural brain images such as magnetic resonance imaging (MRI). How to utilize the multivariate longitudinal outcomes and MRI data to make efficient statistical inference and prediction is an open question. In this article, we propose a multivariate joint model with functional data (MJM-FD) framework that relates multiple correlated longitudinal outcomes to a survival outcome, and use the scalar-on-function regression method to include voxel-based whole-brain MRI data as functional predictors in both longitudinal and survival models. We adopt a Bayesian paradigm to make statistical inference and develop a dynamic prediction framework to predict an individual's future longitudinal outcomes and risk of a survival event. We validate the MJM-FD framework through extensive simulation studies and apply it to the motivating Alzheimer's Disease Neuroimaging Initiative (ADNI) study.

KEYWORDS
ADNI study, Bayesian inference, dynamic prediction, functional regression, multivariate joint model with functional data

1 | INTRODUCTION

Alzheimer's disease (AD) is a common and severe progressive neurodegenerative disorder. In 2020, the predicted number of people in the United States with AD increases to about 5.8 million, compared with around 4.7 million in 2010,[1] which poses a challenging public health issue. To slow the progression of AD, a number of recent studies suggested to evaluate the severity of AD via cognitive and behavioral scores, and to take interventions promptly at early disease stages,[2,3] where a meta-analysis shows that 33.6% mild cognitive impairment (MCI) patients transit to AD after 5 years follow-up.[4] Due

to the complex and heterogeneous nature of AD, a single longitudinal outcome is insufficient to assess AD progression because of the complex pathological mechanisms of AD.[3] Hence, AD studies collect multiple longitudinal neuropsychological and behavioral measurements to assess AD progression, and novel statistical methods are needed to leverage this information.

The joint model (JM) framework is traditionally used to model a longitudinal outcome and a survival outcome.[5,6] In recent biomedical research, multiple longitudinal outcomes are collected for better understanding of complex diseases such as AD, and resulting in more accurate prediction of disease progression.[3] To this end, recent studies extend the JM framework to multivariate settings, where the correlation of multiple longitudinal outcomes are commonly modeled using the shared random effects, and some functional forms of the longitudinal outcomes are included in the survival model.[7-9] A good review of the joint model framework is available in recent literatures.[10,11] One important application of joint model is dynamic prediction,[12-15] where the predicted longitudinal trajectories and survival probabilities are updated in a dynamic fashion for a new individual at different visit times and prediction windows.[11] To the best of our knowledge, no work has been done in incorporating functional data in the multivariate joint model (MJM). In this article, we extend the MJM framework to include high-dimensional functional data as predictors for both longitudinal and survival outcomes. We also propose a personalized dynamic prediction method to predict an individual's future longitudinal outcomes and risk of an event under this framework.

Magnetic resonance imaging (MRI) technique has been widely adopted to detect the volume atrophy for AD patients.[16] Specifically, the structural MRI (sMRI) tool detects the structure, shape, and size of the whole brain, in addition to regions of interest (ROI), which are 3D images consisting of thousands of volumetric voxels.[17] There are two common methods to analyze the sMRI data, ROI-based method and voxel-based method.[18] The ROI-based method combines the volume information in the voxels within a specific region of interest,[19] which may result in losing volumetric information of each voxel. The voxel-based method overcomes the problem of information loss by normalizing and analyzing each voxel separately.[18] For example, a recent voxel-based analysis shows that atrophy in medial and lateral temporal lobe is strongly associated with deterioration of memory scores.[20] However, the voxel-based analysis has other issues such as multiple comparison and failure to account for voxelwise correlation.[21] Because the volumes of some voxels in the whole brain are crucial predictors for AD progression, we adopt the voxel-based whole-brain sMRI data as functional data and consider them as functional predictors for both longitudinal and survival outcomes.

A common approach to model the functional data and handle their high dimensionality is to decompose functions into infinite-dimensional linear combinations of basis functions using basis expansion.[22,23] Then regularization, usually via truncation or penalization, is applied to convert the infinite-dimensional linear combinations to finite linear combinations.[24,25] The scalar-on-function regression is a widely used functional regression method which quantifies the association between a scalar outcome and a functional predictor via the coefficient function.[26] Functional principal component analysis (FPCA) is a powerful dimension reduction approach to decompose the functional predictor,[22,27-30] and splines decomposition has been widely adopted to estimate the coefficient function.[22,23] In addition, the scalar-on-function regression approach has also been developed in the JM framework as functional joint model (FJM).[31]

In this article, we propose a multivariate joint model with functional data (MJM-FD) framework to assess AD progression among MCI patients using baseline characteristics, whole-brain MRI data, and three longitudinal neuropsychological measurements. Comparing with the existing literatures, there are three major novelties in our method. First, we extend FJM by jointly modeling multivariate longitudinal outcomes and a survival outcome, and incorporating functional predictors in the submodels for both types of outcomes. Second, we develop a dynamic prediction framework to predict the future longitudinal trajectories and probability of dementia diagnosis. Third, we use a voxel-selection method to filter informative voxels from whole-brain MRI data.

The article is organized as follows. In Section 2, we introduce the motivating ADNI study, longitudinal outcomes, and MRI data. In Section 3, we propose the multivariate joint model with functional data (MJM-FD), Bayesian inference method, and dynamic prediction framework. In Section 4, we present the results of model fitting and predictive performance in the ADNI study. In Section 5, we conduct an extensive simulation study with eight settings to validate our methods. In Section 6, we summarize our findings and discuss limitations and future work.

2 | A MOTIVATING CLINICAL STUDY

The multivariate joint model with functional data (MJM-FD) framework is motivated by the Alzheimer's Disease Neuroimaging Initiative (ADNI) study, which is a comprehensive longitudinal study to measure the association between

the progression of AD and biomarkers, neuropsychological scores, medical imaging data, and genetic variants and the study consists of four stages: ADNI-1; ADNI-GO; ADNI-2; ADNI-3 (ongoing stage). We downloaded all clinical and MRI data on 02 November 2019. In summary, the total number of subjects is 2253, with 815 cognitive normal (CN), 1018 MCI, 390 dementia participants, and 30 participants with missing diagnosis at baseline. The mean follow-up time is 2.43 years (a mean of 6.46 visits). More information for the ADNI study can be found on the ADNI website: http://adni.loni.usc. edu/. We include all patients with MCI diagnosis and valid baseline MRI data, APOE-ϵ4, and age. The final analysis dataset consists of 742 patients, with a total of 3826 longitudinal observations before dementia diagnosis or censoring, a mean follow-up length of 3.10 years, and standard deviation of 2.89 years.

We select three longitudinal outcomes: Mini-Mental State Examination (MMSE; score ranges from 0 to 30 with higher score indicating better cognitive function), clinical dementia rating—sum of boxes (CDR-SB; score ranges from 0 to 18 with higher score indicating worse cognitive behavior), Wechsler memory scale— logical memory (WMS-LM; score ranges from 0 to 25 with higher score indicating better memory function) due to two reasons. First, MMSE and CDR-SB are reported to be significantly associated with AD progression.[32] Second, all three measures are commonly collected in AD studies.[33-35] Specifically, MMSE is a widely adopted scale to measure the cognitive function and is closely related to AD progression.[36,37] CDR-SB evaluates the cognitive functioning in six domains (memory, orientation, judgment and problem solving, community affairs, home and hobbies, and personal care) and it is closely associated with AD progression.[36] Moreover, WMS-LM is a measurement of cognitive function in the memory aspect, and the logical memory subset of WMS is a good measure in distinguishing between subjects who develop AD and who remain stable.[37]

The ADNI study has different schedule for MMSE, CDR-SB, and WMS-LM. For MCI patients, MMSE and CDR-SB are measured at baseline, month 6 (ADNI-1, ADNI-GO, ADNI-2 phases only), month 12, month 18 (ADNI-1 phase only), month 24 and annual visits after, while WMS-LM are measured at baseline, month 12, and annual visits after. Thus, we obtain the analysis dataset with 3593 visits (89 missing), 3593 visits (82 missing), 2771 visits (92 missing) for MMSE, CDR-SB, and WMS-LM, respectively, before dementia diagnosis or censoring. Figure 1 displays the longitudinal trajectories in LOWESS curves of three outcomes using visits at or before dementia diagnosis, which suggests nonlinear trajectory patterns. The missing longitudinal data are multiply imputed and averaged using bootstrap EM algorithm via Amelia package. The multiple imputation model includes baseline covariates (age, gender, education years, and APOE-ϵ4 alleles), other two longitudinal scores, and current diagnosis status (CN, MCI or dementia).

We use the baseline voxel-based whole-brain MRI data as functional predictors, where the MRI data go through correction steps before preprocessing. More details about correction steps and MRI preprocessing protocols are given in Section 3.5.
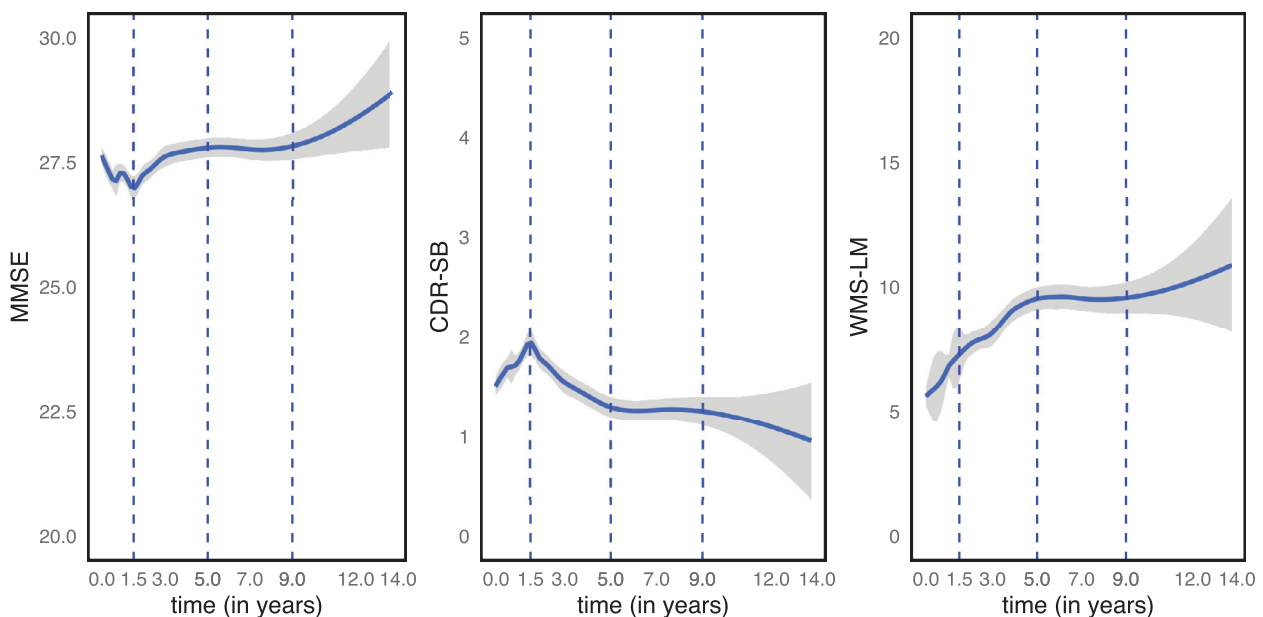

FIGURE 1 The longitudinal trajectories of MMSE, CDR-SB, and WMS-LM for visits at or before dementia diagnosis. The solid curves are the LOWESS curves with 95% CI (shaded regions). The dashed vertical lines are visit times at 1.5, 5.0, and 9.0 years. The slopes are markedly different at those time points [Colour figure can be viewed at wileyonlinelibrary.com]

3 | METHODS

In brief, our multivariate joint model with functional data (MJM-FD) consists of three major steps. First, we select informative voxels in structural MRI data using linear mixed models and Cox proportional hazards model to reduce the amount of MRI data, and we transform the volumes of selected voxels into functional predictors. Second, functional principal component analysis and cubic B-spline approximation are adopted to reduce dimensions of functional predictors and coefficient functions, respectively. Third, the joint modeling of longitudinal and survival data are incorporated in the MJM-FD model and Bayesian methods are used for statistical inference.

3.1 | Voxel selection

After preprocessing steps detailed in Section 3.5, sMRI data have the same image structure and format as the Eve Template, with $S = 1\,843\,303$ non-zero-volume voxels. To reduce the computation time and to resolve the large memory issue associated with the high-dimensional functional data, we adopt the false discovery rate (FDR) adjustment to select informative voxels to be included in the real data analysis.

The voxel-selection step we adopt is similar, in principle, to voxel-based morphometry (VBM), which is an automated procedure to detect differences in voxel volumes between groups using voxelwise comparison method.[38,39] The VBM method has been widely used to assess the association between voxel volumes and cognitive change, both cross-sectionally and longitudinally, while adjusting for other covariates, such as age, sex, education, and so on.[20,40-44] In these analyses, the association between cognitive scores and voxel volumes is commonly assessed using the regression methods, but the interaction term between voxel volume and time is typically not included as a regression covariate.[45] Thus, in the voxel-selection step, we do not include the interaction between the voxel volume and time.

We have $i = 1, 2, \ldots, n$ subjects, $j = 1, 2, \ldots, J$ longitudinal measurements, and $k = 1, 2, \ldots K_{ij}$ visits for subject i and outcome j. We denote the longitudinal measurements $\{Y_{ij}(t_{ijk})\}$ and $\{\tilde{Y}_{ij}(t_{ijk})\}$ at time t_{ijk} for observed and missing longitudinal outcomes, respectively. We assume missing at random (MAR) and denote the missing indicator as I_{ijk} (1 if missing and 0 otherwise). We denote p_j-dimensional covariate vector as $\boldsymbol{x_{ijk}}$, the event time $T_i = \min(T_i^*, C_i)$, with T_i^* being the failure time and C_i being the censoring time. We observe the censoring indicator $\delta_i = I(T_i^* \leq C_i)$, and the p_w-dimensional covariate vector $\boldsymbol{w_i}$. For subject i, the MRI data $g_i(\boldsymbol{S})$ is defined on a dense domain $\boldsymbol{S}$ with S voxels, and the volume of voxel s is denoted by $\text{Vol}_{i,s}$, $s = 1, 2, \ldots, S$.

For each longitudinal outcome j, we independently fit S linear mixed effects models $Y_{ij}(t_{ijk}) = \beta_{0j} + \boldsymbol{x'_{ijk}}\boldsymbol{\beta_j} + \text{Vol}_{i,s}\beta_{js} + u_{ij0} + t_{ijk}u_{ij1} + \epsilon_{ijk}$, where β_{0j} is the intercept, $\boldsymbol{\beta_j}$ is the p_j-dimensional regression coefficient, β_{js} is the regression coefficient for voxel s. We assume that the random intercepts $u_{ij0} \sim N(0, \sigma_{uj0}^2)$, and random slopes $u_{ij1} \sim N(0, \sigma_{uj1}^2)$, with correlation coefficient ρ_j, and the random errors $\epsilon_{ijk} \sim N(0, \sigma_j^2)$. For the survival outcome, we independently fit S Cox proportional hazards models $h_i(t) = h_0(t)\exp(\boldsymbol{w'_i}\boldsymbol{\gamma} + \text{Vol}_{i,s}\gamma_s)$, where $h_0(t)$ is the baseline hazard function, $\boldsymbol{\gamma}$ is the p_w-dimensional regression coefficient vector, and γ_s is the regression coefficient for voxel s.

The missing longitudinal outcome data are imputed using bootstrap EM algorithm as in Section 2, generating five imputed datasets. For each imputed dataset, we use FDR to adjust the S p-values of the regression coefficients of voxel volumes for longitudinal outcomes, and we take union of the voxels with adjusted p-values of regression coefficients less than the FDR threshold. We use FDR-adjusted p-value threshold of 0.05 for MMSE and CDR-SB, and a more conservative threshold of 0.01 for WMS-LM and survival outcome. We select more stringent threshold for WMS-LM and the survival outcome, because we would get more than 10 000 voxels with 0.05 being the threshold, which requires intensive computation time and memory. After voxel selection, we obtain functional predictor $g_{i,j}(s)$ defined on domain $\boldsymbol{S_j}$ for each longitudinal outcome j and functional predictor $g_{i,w}(s)$ defined on domain $\boldsymbol{S_w}$ for the survival outcome.

3.2 | Multivariate joint model with functional data

We propose the MJM-FD consisting of two submodels:

$$Y_{ij}(t_{ijk}) = m_{ij}(t_{ijk}) + \epsilon_{ijk}, \tag{1}$$

$$\text{where} \quad m_{ij}(t_{ijk}) = \beta_{0j} + \boldsymbol{x}'_{ijk}\boldsymbol{\beta_j} + \int_S g_{i,j}(s)B_j(s)ds + v_{j0}u_{i0} + v_{j1}t_{ijk}u_{i1}, \tag{2}$$

$$h_i(t) = h_0(t)\exp\left(\boldsymbol{w}'_i\boldsymbol{\gamma} + \int_S g_{i,w}(s)B_w(s)ds + \sum_{j=1}^J \alpha_j m_{ij}(t)\right). \tag{3}$$

The latent longitudinal means are denoted as $m_{ij}(t)$ for subject i and longitudinal outcome j at visit time t. We denote the coefficient functions as $B_j(s)$ for the longitudinal outcome j, and $B_w(s)$ for the survival outcome. For simplicity and computational efficiency, we model the correlation among longitudinal outcomes via the shared random effects $\boldsymbol{u}'_i = (u_{i0}, u_{i1})'$, with outcome-specific association parameters v_{j0} and v_{j1} for $j = 1, 2, \ldots, J$, while other forms of multivariate random effects can be readily adopted.[46] We assume that the random intercepts $u_{i0} \sim N(0, \sigma^2_{u0})$, and random slopes $u_{i1} \sim N(0, \sigma^2_{u1})$ with correlation coefficient ρ, and are shared across all J longitudinal outcomes. We let the association parameter α_j denote the association between the survival outcome and longitudinal outcome j. We set $v_{10} = v_{11} = 1$ for model identifiability. To account for the skewness of longitudinal outcomes, we let the random errors $\epsilon_{ijk} \sim$ Skew-Normal$(0, \omega_j^2, \delta_j)$, where ω_j is the scale parameter and δ_j is the shape parameter (reduces to Normal distribution if $\delta_j = 0$). We approximate the baseline hazard function $h_0(t)$ using piecewise constant functions.[47-49] Specifically, we set $h_0(t) = h_{0z}$ if $\tau_z \leq t < \tau_{z+1}$ for knot locations $\boldsymbol{\tau} = (\tau_1, \ldots, \tau_Z)$, where $\tau_1 = 0$.

We adopt functional principal component analysis (FPCA) to decompose the functional predictors $g_{i,j}(s)$ and $g_{i,w}(s)$, and adopt B-spline to decompose the coefficient functions $B_j(s)$ and $B_w(s)$. For notational convenience, we denote the functional predictor as $g_i(s)$ and its corresponding coefficient function as $B(s)$. The functional predictor is observed as a noisy form: $g'_i(s) = g_i(s) + \epsilon_i(s)$, where $\epsilon_i(s)$ is a stochastic Gaussian process with mean zero, and variance σ^2. We define the mean function $\mu(s) = E(g_i(s))$, and the covariance function $C(s, s') = \text{cov}(g_i(s), g_i(s'))$, where s and s' denote two points over domain S. By Mercer's theorem, $C(s, s') = \sum_{l=1}^\infty a_l\phi_l(s)\phi_l(s')$, where $a_1 \geq a_2 \geq \ldots 0$ are non-increasing eigenvalues, and $\phi_l(s)$ is the corresponding eigenfunction, satisfying $\int_S \phi_l(s)^2 ds = 1$ and $\int_S \phi_l(s)\phi_{l'}(s)ds = 0$ when $l \neq l'$. We decompose the functional predictor $g_i(s) = \mu(s) + \sum_{l=1}^\infty \xi_{il}\phi_l(s)$ using Karhunen-Loeve's expansion, where ξ_{il} is the corresponding FPC score, satisfying $E(\xi_{il}) = 0, \text{Var}(\xi_{il}) = a_l, \xi_{il} \perp\!\!\!\perp \xi_{il'}$ for $l \neq l'$, where $\perp\!\!\!\perp$ denotes independence. The truncated Karhunen-Loeve's expansion is adopted, using proportion of variance explained (PVE) criteria: $\sum_{l=1}^L a_l / \sum_{l=1}^\infty a_l \geq$ PVE.

We conduct FPCA using the following procedure for the observed $g'_i(s)$. First, the estimated mean function $\hat{\mu}(s)$ is obtained from the sample mean: $\hat{\mu}(s) = (1/n)\sum_{i=1}^n g'_i(s)$, and is smoothed using the cubic B-spline function. Second, we demean the data $g'_i(s) - \hat{\mu}(s)$ from which we calculate the raw covariance matrix. Third, the raw covariance matrix is smoothed using either the tensor-product smoother (for low-dimensional domains), or a bivariate scattered points from the multilevel B-splines (for high-dimensional domains). Finally, we perform eigendecomposition for the smoothed covariance matrix to obtain the estimated eigenfunctions $\hat{\phi}_l(s)$ and the number of eigenfunctions L, where L is set to be the minimum number that achieves the PVE criteria.

For the coefficient function $B(s)$, we decompose it using cubic B-spline functions: $B(s) = \sum_{p=1}^P b_p f_p(s)$, where $f_p(s)$ is the cubic B-spline function p evaluated at location s, and b_p is the corresponding B-spline coefficient. For model identifiability, the number of B-spline functions P is smaller than or equal to the number of eigenfunctions L. Hence, we set $P = \min(L_1, \ldots, L_J, L_w)$.

We approximate the integration term $\int_S g_i(s)B(s)ds$ as $\int_S g_i(s)B(s)ds \approx \boldsymbol{I}_{\mu B}\boldsymbol{b} + \boldsymbol{\Xi}_i\boldsymbol{I}_{\phi B}\boldsymbol{b}$, where $\boldsymbol{I}_{\mu B}$ is a 1 by P vector, with pth element as $\int_S \mu(s)f_p(s)ds$, and $\boldsymbol{b}$ is a P by 1 coefficient vector, with pth element as b_p. We denote $\boldsymbol{\Xi}_i$ as a 1 by L vector, with lth element as ξ_{il}, and $\boldsymbol{I}_{\phi B}$ as an L by P matrix, with (l, p)th element being $\int_S \phi_l(s)f_p(s)ds$ (refer to Web Supplement Section 1 for derivation). The integrals $\int_S \mu(s)f_p(s)ds$ and $\int_S \phi_l(s)f_p(s)ds$ are approximated using Riemann sum. The FPC scores satisfy $\xi_{il} = \int_S (g_i(s) - \mu(s))\phi_l(s)ds$, which are approximated as $\hat{\xi}_{il} \approx \int_S (g'_i(s) - \hat{\mu}(s))\hat{\phi}_l(s)ds$.

3.3 | Estimation and inference

We denote the parameter space as: $\boldsymbol{\theta} = (\boldsymbol{\beta}_0, \boldsymbol{\beta}, \boldsymbol{b}, \boldsymbol{\sigma_u}, \rho, \boldsymbol{v}, \boldsymbol{\omega}, \boldsymbol{\delta}, \boldsymbol{h}_0, \boldsymbol{\gamma}, \boldsymbol{\alpha})'$, where $\boldsymbol{\beta}_0 = (\beta_{01}, \ldots, \beta_{0J})$, $\boldsymbol{\beta} = (\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_J)$, $\boldsymbol{b} = (\boldsymbol{b}_1, \ldots, \boldsymbol{b}_J, \boldsymbol{b}_w)$, $\boldsymbol{b}_j = (b_{j1}, \ldots, b_{jp})$, $\boldsymbol{b}_w = (b_{w1}, \ldots, b_{wp})$, $\boldsymbol{\sigma_u} = (\sigma_{u0}, \sigma_{u1})$, $\boldsymbol{v} = (v_{20}, v_{21}, \ldots, v_{J0}, v_{J1})$, $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_J)$, $\boldsymbol{\delta} = (\delta_1, \ldots, \delta_J)$, $\boldsymbol{h}_0 = (h_{01}, \ldots, h_{0Z})$, and $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_J)$.

We express the full data log-likelihood as:

$$l(\boldsymbol{\theta}|Y, \tilde{Y}, T, \delta) = \sum_{i=1}^{n} \log \left\{ \left[\prod_{j=1}^{J} \prod_{k=1}^{K_{ij}} f(y_{ij}(t_{ijk})|\boldsymbol{u_i}, \boldsymbol{\theta})^{1-I_{ijk}} f(\tilde{y}_{ij}(t_{ijk})|\boldsymbol{u_i}, \boldsymbol{\theta})^{I_{ijk}} \right] \cdot \right.$$
$$\left. h_i(T_i|\boldsymbol{u_i}, \boldsymbol{\theta})^{\delta_i} S_i(T_i|\boldsymbol{u_i}, \boldsymbol{\theta}) P(\boldsymbol{u_i}|\boldsymbol{\theta}) \right\}, \tag{4}$$

where $\tilde{y}_{ij}(t_{ijk})|\boldsymbol{u_i}, \boldsymbol{\theta} \sim$ Skew-Normal$(m_{ij}(t_{ijk}), \omega_j^2, \delta_j)$, and $S_i(T_i)$ is the survival function. The missing outcomes $\tilde{y}_{ij}(t_{ijk})$ are imputed in the MCMC sampling procedure. Specifically, we treat $\tilde{y}_{ij}(t_{ijk})$ as unknown parameters and impute from the Skew-Normal distribution with each parameter being the posterior samples.

To sample from the log-likelihood, we utilize the software Stan, which provides a platform to perform full Bayesian inference using no-u-turn sampler (NUTS).[50] The NUTS algorithm performs more efficiently and faster than the Hamiltonian Monte Carlo (HMC) sampler. In our real data application, we use two Monte Carlo chains with overdispersed initial values and each chain has 3000 iterations with 2000 iterations as burn-in iterations. We use Gelman-Rubin diagnostic to ensure the scale reduction $\hat{R}$ of all parameters are smaller than 1.1 as well as a suite of convergence diagnosis criteria to ensure convergence.[50] The results consist of 2000 NUTS samples for each parameter, from which we derive the posterior mean, estimated standard error, and the 95% credible interval for each parameter. The Stan code is provided in the Web Supplement Section 8 for easy implementation.

To evaluate model performance, we adopt leave-one-out cross-validation information criteria (LOOIC) and widely applicable information criteria (WAIC), which are robust, stable, and computationally efficient in Bayesian settings.[51] Smaller LOOIC and WAIC values indicate better model performance.

3.4 | Dynamic prediction

Suppose we have a new patient M who is event-free to time T_0. We predict future longitudinal outcomes at time T_1, $T_1 > T_0$, and the conditional survival probability. We use NUTS to obtain the samples of the random effects $\boldsymbol{u_M}$ from its posterior distribution (refer to the Web Supplement Sections 2 and 3 for the derivation of posterior distributions of random effects and conditional survival probability).

$$P\left(\boldsymbol{u_M}|T_M > T_0, \boldsymbol{y}_M, \hat{\boldsymbol{\theta}}\right) \propto \prod_{j=1}^{J} \prod_{k=1}^{K_{Mj}} P\left(y_{Mj}(t_{Mjk})|\boldsymbol{u_M}, T_M > T_0, \hat{\boldsymbol{\theta}}\right) P\left(T_M > T_0|\boldsymbol{u_M}, \hat{\boldsymbol{\theta}}\right) P\left(\boldsymbol{u_M}|\hat{\boldsymbol{\theta}}\right),$$

where we denote $\boldsymbol{y}_M = (y_{M1}(t_{11}), \ldots, y_{M1}(T_0), \ldots, y_{MJ}(t_{J1}), \ldots, y_{MJ}(T_0))$ as all longitudinal outcomes up to time T_0 for subject M, where the subscript M denotes subject M, and K_{Mj} is the number of time points up to time T for subject M, outcome j.

After drawing D samples of random effects $\boldsymbol{u_M}^{(d)}, d = 1, 2, \ldots, D$, the latent longitudinal mean for the longitudinal outcome j at time T_1 and conditional survival probability are calculated as:

$$E\left(y_{Mj}^{(d)}(T_1)|T_M > T_0, \boldsymbol{y}_M, \boldsymbol{u_M}^{(d)}, \hat{\boldsymbol{\theta}}\right) = \hat{\beta}_{0j} + \boldsymbol{x}'_{Mj}(T_1)\hat{\boldsymbol{\beta}}_j + \boldsymbol{B}'_{Mj}\hat{\boldsymbol{b}}_j + \hat{v}_{j0}u_{M0}^{(d)} + \hat{v}_{j1}T_1 u_{M1}^{(d)},$$

$$P\left(T_M > T_1|T_M > T_0, \boldsymbol{y}_M, \boldsymbol{u_M}^{(d)}, \hat{\boldsymbol{\theta}}\right) = \exp\left(-\int_{t=T_0}^{T_1} h_M\left(t|\boldsymbol{y}_M, \boldsymbol{u_M}^{(d)}, \hat{\boldsymbol{\theta}}\right) dt\right),$$

where $h_M(t|\boldsymbol{y}_M, \boldsymbol{u_M}^{(d)}, \hat{\boldsymbol{\theta}}) = \hat{h}_0(t) \exp(\boldsymbol{w}'_M \hat{\boldsymbol{\gamma}} + \boldsymbol{B}'_{Mw} \hat{\boldsymbol{b}}_w + \sum_{j=1}^{J} \hat{\alpha}_j E(y_{Mj}^{(d)}(t)))$. We denote $\boldsymbol{x}'_{Mj}(T_1)$ as the observed longitudinal covariate vector at time T_1 for longitudinal outcome j, $\boldsymbol{w}'_M$ as the observed survival covariate vector, $\boldsymbol{B}'_{Mj}$ and $\boldsymbol{B}'_{Mw}$ as the functional covariate vectors, where $\boldsymbol{B}'_{Mj} = \boldsymbol{I}_{\mu_j, B_j} + \boldsymbol{\Xi}_{Mj} \boldsymbol{I}_{\phi_j, B_j}$, and $\boldsymbol{B}'_{Mw}$ is calculated similarly. The FPC score vector $\boldsymbol{\Xi}_{Mj}$ is approximated as in Section 3.2. The means and quantiles of longitudinal means and conditional survival probability are calculated based on the D samples.

We evaluate the predictive performance in discrimination (how well the models discriminate between subjects who had the event from those who did not) and calibration (how well the models predict the observed data). We assess

discrimination using time-dependent area under the receiver operating characteristic curve (AUC). Specifically, we adopt an efficient and unbiased estimator for time-dependent sensitivity and specificity proposed by Li et al.[52] The time-dependent sensitivity is estimated as $\hat{P}(X_i > c | T_i \leq \tau) = \sum_{i=1}^{n} W_i I(X_i > c) / \sum_{i=1}^{n} W_i$, and the time-dependent specificity is estimated as $\hat{P}(X_i \leq c | T_i \geq \tau) = \sum_{i=1}^{n}(1 - W_i)I(X_i \leq c) / \sum_{i=1}^{n}(1 - W_i)$, where W_i is the probability of subject i not surviving to time t. The AUC can be approximated using trapezoidal integration from the time-dependent ROC curve, where the best AUC value is 1, indicating perfect prediction, and an AUC value of 0.5 indicates no better than a random guess. The calibration is assessed by Brier score (BS), which measures the disparity of prediction at a particular time, using a weighted sum of squares of the difference between the true survival status and the estimated survival probability.[53] The best BS value is 0, and a BS value 0.25 indicates no better than a random guess.

3.5 | MRI preprocess protocols and software implementation

The downloaded sMRI data had been corrected using gradient warping, B1 nonuniformity correction, and N3 correction. We utilize the MRI preprocess pipelines from Advanced Normalization Tools (ANTs) and FMRIB Software Library (FSL) to process the corrected 3D sMRI data in the following steps: (1) The sMRI data are stored as Digital Imaging and Communications in Medicine (DICOM) format, which are transformed into Neuroimaging Informatics Technology Initiative (NIFTI) format, using `oro.dicom` package v0.5.3 and `oro.nifti` package v0.10.1. (2) The neck part in the image is removed using standard `MNI152 T1 1mm` brain template and `extrantsr` package v3.9.8.1, and the skull part is then stripped using `fslr` package v2.24.2. (3) The neck-skull-stripped image is registered to the JHU-MNI-ss (Eve) Template,[54] using Symmetric Normalization (SyN) transformation, and Lanczos Windowed Sinc interpolator via `extrantsr` package and `EveTemplate` package v0.99.14. (4) The registered image is standardized to have mean 0 and standard deviation 1, and the image voxels are extracted and merged with the White Matter Parcellation Map (WMPM Type II) of the `Eve Template`,[55] using `neurobase` package v1.27.11.2 and `EveTemplate` package. (5) We align the non-zero-volume voxels in the order of three-dimensional coordinates, which are then transformed to the one-dimensional domain.

We utilize the following software packages for other computational steps. We use packages `mgcv` v1.8-30 and `MBA` v0.0-9 for the tensor-product smoother, and bivariate scattered points from the multilevel B-splines, respectively. We use package `Rspectra` v0.16-0 for high-dimensional eigendecomposition and we obtain the values of the B-spline function using `spline` package v4.0.4. For Bayesian inference, we use `rstan` package v2.19.3. We use `loo` package v2.1.0 to compute LOOIC and WAIC. For the estimation of AUC and BS, we use package `tdROC` v1.0 and `ipred` v0.0-9, respectively.

In summary, our `rstan` code can handle the following scenarios: (1) shared random effects; (2) more than three longitudinal outcomes by specifying the likelihood for each longitudinal outcome; (3) monotone or nonmonotone missing outcomes under missing at random (MAR) assumption; (4) our MJM-FD model reduces to multivariate joint models by removing the functional predictors and coefficient functions.

4 | APPLICATION TO THE ADNI STUDY

4.1 | Model description and performance

The baseline characteristics we include are age (mean = 72.8, sd = 7.7), gender (41.4% female), education years (mean = 16.0, sd = 2.8), Apolipoprotein E (APOE)-ϵ4 alleles (51.2%, 38.1%, 10.6% patients with 0, 1, 2 alleles, respectively). We propose the MJM-FD:

$$Y_{ij}\left(t_{ijk}\right) = m_{ij}\left(t_{ijk}\right) + \epsilon_{ijk},$$

$$\text{where } m_{ij}\left(t_{ijk}\right) = \beta_{0j} + \sum_{\kappa=1}^{4}\left(t_{ijk} - \tau_\kappa\right)_{+}\beta_{1j\kappa} + \boldsymbol{x}_i'\boldsymbol{\beta_j} + \int_S g_{i,j}(s)B_j(s)ds + v_{j0}u_{i0} + v_{j1}t_{ijk}u_{i1},$$

$$h_i(t) = h_0(t)\exp\left(\boldsymbol{w}_i'\boldsymbol{\gamma} + \int_S g_{i,w}(s)B_w(s)ds + \sum_{j=1}^{J}\alpha_j m_{ij}(t)\right),$$

where $(t_{ijk} - \tau_\kappa)_+$ (equal to $t_{ijk} - \tau_\kappa$ if $t_{ijk} > \tau_\kappa$ and 0 otherwise) are spline terms to account for the nonlinearity of longitudinal trajectories in Figure 1, while $\tau = \{\tau_\kappa\}_{\kappa=1,\ldots,4} = (0, 1.5, 5.0, 9.0)$ are knot locations. Let $Y_{i1}(t_{ijk}), Y_{i2}(t_{ijk}), Y_{i3}(t_{ijk})$ be the observed longitudinal outcomes MMSE, CDR-SB, WMS-LM, respectively, for subject i at visit time t_{ijk} from baseline. The survival time is the time from baseline to dementia diagnosis or censoring. Note that death is not reported in the ADNI study and we have no information after each individual's last visit. If dementia diagnosis is not made during the study, we set the dementia diagnosis status as censoring at the last visit. The baseline hazard function is modeled using a piecewise constant function, using same knot locations as the longitudinal time spline. We set baseline characteristics $\boldsymbol{x}'_i = (\text{Age}_i, \text{Gender}_i, \text{Education}_i)'$, and $\boldsymbol{w}'_i = (\text{Age}_i, \text{Gender}_i, \text{Education}_i, \text{APOE}_i)'$, where APOE_i is the number of APOE-ϵ4 alleles for subject i. We follow the preprocess protocols and voxel-selection methods to process the baseline MRI data. The domains $\boldsymbol{S_1}, \boldsymbol{S_2}, \boldsymbol{S_3}$, and $\boldsymbol{S_w}$ consist of 324, 69, 1387, and 972 voxels, respectively.

We compare model performance of six models: MJM-FD (defined above), Hippocampal Model (HM; MJM-FD using only hippocampal voxels as functional predictors $g_i(s)$ for both longitudinal and survival outcomes, with $g_i(s)$ on domain $\boldsymbol{S}$ with 9846 voxels), Functional Cox model (FCM, $\alpha_1 = \alpha_2 = \alpha_3 = 0$ in MJM-FD), baseline functional Cox model (BFCM, FCM with longitudinal outcome data at baseline), Multivariate joint model (MJM; MJM-FD without functional predictors), and MJM-FCM (MJM-FD without functional predictors in longitudinal models). For FCM and BFCM, the parameters for longitudinal models are also estimated and the log-likelihood calculation also includes the longitudinal model, in order to have a fair comparison of LOOIC and WAIC. The reason why we compare MJM-FD and HM is that atrophy of hippocampus volume is strongly associated with faster progression from MCI to AD,[56] and we desire to compare the performance of models using voxels from whole-brain vs hippocampus only.

For model MJM-FD, we perform FPCA on the observed functional predictors and retain certain numbers of eigenfunctions (8 for MMSE, 6 for CDR-SB, 7 for WMS-LM, and 7 for the survival outcome), with PVEs all above 99%. Given the small numbers of informative voxels in each outcome, we use tensor-product surface spline to smooth the raw covariance matrices. For model HM, we retain the first 10 eigenfunctions, with PVE above 95%, and we use bivariate scattered points from the multilevel B-splines to smooth the covariance matrices. Next, we perform eigendecomposition for the smoothed covariance matrices to obtain estimated eigenfunctions and eigenvalues, and construct the approximated covariate matrices.

Table 1 displays leave-one-out cross-validation information criteria (LOOIC) and widely applicable information criteria (WAIC) from six candidate models. Model MJM-FD has smaller LOOIC and WAIC compared with HM, indicating that the inclusion of whole-brain voxels improves model fit performances compared with hippocampus voxels only. Comparing with FCM and BFCM, MJM-FD shows large improvement of LOOIC and WAIC, suggesting the advantage of including longitudinal trajectories in the survival model. Moreover, MJM-FD outperforms MJM and MJM-FCM, suggesting that functional predictors are predictive to both longitudinal and survival modeling. The proposed model MJM-FD performed markedly better than all other candidate models with smallest LOOIC and WAIC. Hence, we select model MJM-FD as the final model.

4.2 | Results for model fitting

The parameter estimates from model MJM-FD are presented in Table 2. For the outcome Mini-Mental State Examination (MMSE), the average decrease is 0.164 (95% CI: [−0.298, −0.035]) for females compared with males, and the average increase is 0.070 (95% CI: [0.046, 0.095]) for 1 year increase in education, conditioning on other covariates, and the random effects. For the outcome Wechsler Memory Scale—logical memory (WMS-LM), the average decrease is 0.037 (95% CI: [−0.076, −0.001]) for 1 year increase in baseline age, and the average increase is 0.374 (95% CI: [0.283, 0.463]) for

TABLE 1 Leave-one-out cross-validation information criteria (LOOIC) and widely applicable information criteria (WAIC) from MJM-FD, hippocampal model (HM), functional Cox model (FCM), baseline functional Cox model (BFCM), multivariate joint model (MJM), and MJM-FCM

	MJM-FD	**HM**	**FCM**	**BFCM**	**MJM**	**MJM-FCM**
LOOIC	37 875.4	38 007.3	38 252.2	38 166.7	37 943.3	37 948.8
WAIC	37 395.3	37 528.4	37 791.7	37 706.6	37 445.0	37 470.0

TABLE 2 Estimated mean, standard error, and 95% credible intervals from model MJM-FD

	Parameters	Mean	SE	2.5%	97.5%
Longitudinal outcomes					
MMSE	Intercept	29.055	0.433	28.226	29.873
	Age (years)	−0.006	0.005	−0.016	0.004
	Female sex	−0.164	0.068	−0.298	−0.035
	Education (years)	0.070	0.012	0.046	0.095
CDR-SB	Intercept	0.768	0.174	0.427	1.119
	Age (years)	−0.003	0.002	−0.007	0.001
	Female sex	0.009	0.027	−0.042	0.061
	Education (years)	−0.007	0.005	−0.016	0.002
WMS-LM	Intercept	4.794	2.225	0.436	9.158
	Age (years)	−0.037	0.018	−0.076	−0.001
	Female sex	−0.145	0.261	−0.667	0.388
	Education (years)	0.374	0.046	0.283	0.463
	σ_{u1}	0.557	0.067	0.440	0.695
	σ_{u2}	0.452	0.042	0.370	0.531
	ρ	0.473	0.046	0.375	0.563
	v_{20}	−0.009	0.036	−0.078	0.064
	v_{21}	−1.331	0.109	−1.573	−1.154
	v_{30}	5.573	0.656	4.398	6.936
	v_{31}	2.696	0.231	2.284	3.185
	δ_1	−5.329	0.511	−6.416	−4.401
	δ_2	7.816	0.931	6.288	9.933
	δ_3	0.049	0.565	−0.937	0.962
Survival outcome	Age (years)	−0.016	0.010	−0.034	0.003
	Female sex	−0.083	0.144	−0.376	0.214
	Education (years)	0.006	0.025	−0.044	0.055
	APOE-ϵ4 Alleles	0.285	0.096	0.104	0.469
	α (MMSE)	−0.073	0.173	−0.431	0.253
	α (CDR-SB)	0.300	0.124	0.050	0.540
	α (WMS-LM)	−0.162	0.044	−0.248	−0.075

Note: The estimated coefficients for coefficient functions are available in Web Supplement Section 5.

1 year increase in education, conditioning on other covariates, and the random effects. The random intercepts and the random slopes of time are positively correlated, with a correlation coefficient of 0.473 (95% CI: [0.375, 0.563]), suggesting that MCI patients with worse cognition scores at baseline (negative random intercept u_{i0}) tend to have faster rate of deterioration (negative random slope u_{i1}). The estimated shape parameters are −5.329 (95% CI: [−6.416, −4.401]), 7.816 (95% CI: [6.288, 9.933]), and 0.049 (95% CI: [−0.937, 0.962]) for MMSE, CDR-SB, and WMS-LM, respectively, indicating left-skewed distribution for MMSE, right-skewed distribution for CDR-SB, and no significant skewness for WMS-LM. The estimated association parameters v_{20} and v_{21} are negative, while the estimated association parameters v_{30} and v_{31} are positive, suggesting that higher (better) MMSE scores are associated with lower (better) CDR-SB scores and higher (better) WMS-LM scores. Specifically, one unit increase of random intercepts is associated with 5.573 (95% CI: [6.288, 9.933]) unit
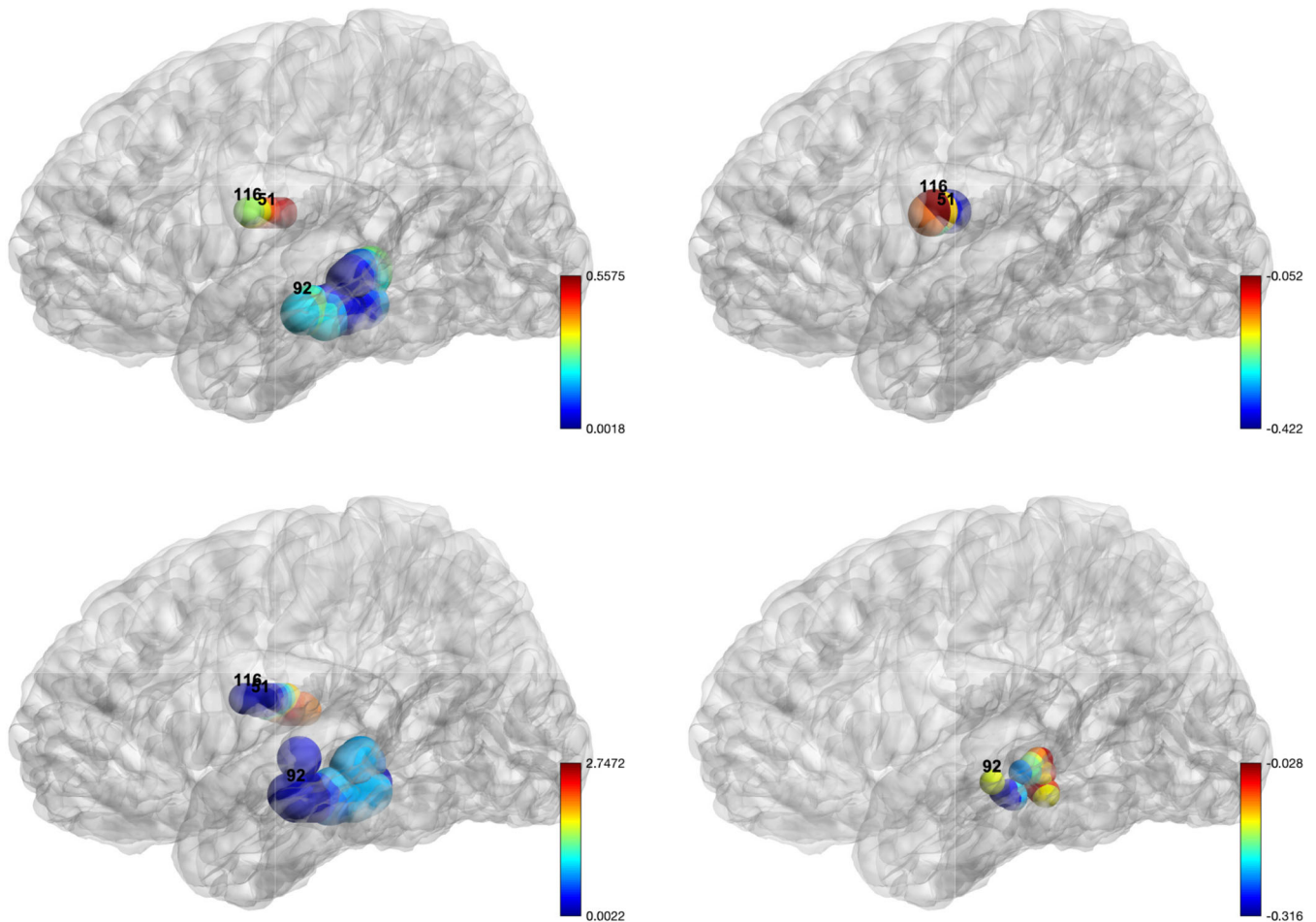
FIGURE 2 The brain (in sagittal view) heat map of estimated coefficient functions. For clarity, we only display voxels in the right hippocampus (label 92), right fornix column and body (label 116), and left fornix column and body (label 51). The top left, top right, bottom left, and bottom right panels are estimated coefficient functions for MMSE, CDR-SB, WMS-LM, and survival outcome, respectively. The colors indicate values of the estimated mean of coefficient functions. The node sizes are proportional to the number of significant voxels in the ROIs for each outcome. The numbers are ROI labels (refer to Supplementary Tables 6 and 7) [Colour figure can be viewed at wileyonlinelibrary.com]

increase of WMS-LM score, while keeping all other covariates as constants. Moreover, the subject-specific progression rates of MMSE and CDR-SB are negatively associated ($\hat{v}_{21} = -1.331$, 95% CI: $[-1.573, -1.154]$), while the subject-specific progression rates of MMSE and WMS-LM are positively associated ($\hat{v}_{31} = 2.696$, 95% CI: $[2.284, 3.185]$).

For the survival outcome, APOE-$\epsilon 4$ is the only significant predictor for the survival outcome. One additional APOE-$\epsilon 4$ allele is associated with an increase of logarithm hazard rate by 28.5% (95% CI: $[0.104, 0.469]$), conditioning on other covariates. One unit increase of CDR-SB score and WMS-LM score is associated with increase of logarithm hazard rate of 30.0% (95% CI: $[0.050, 0.540]$) and the decrease of logarithm hazard rate of 16.2% (95% CI: $[-0.248, -0.075]$), respectively.

Using BrainNet Viewer,[57] we plot in Figure 2 (in sagittal view of the Eve template) the heat map of the estimated coefficient functions. For clearer visualization, we provide the 3D movies of the estimated functional predictors in the Web Supplement Section 7. We select some voxels for illustration purpose. For MMSE, the atrophy of voxels on right and left fornix columns and body (label 116 and label 51), right hippocampus (label 92) is associated with lower MMSE scores (worse cognitive functions). For CDR-SB, the atrophy of voxels on right and left fornix columns and body is associated with higher CDR-SB scores (worse behavioral functions). For WMS-LM, the atrophy of voxels on right and left fornix columns and body, and right hippocampus is associated with lower WMS-LM scores (worse memory functions). For the survival outcome, the atrophy of voxels on right hippocampus is associated with faster conversion from MCI to AD. Previous studies find out that the atrophy of the hippocampus region is associated with worse cognitive functions and faster transition to AD.[31] Our results confirm this finding, and discover some voxels on other ROIs that are associated with cognitive functions and AD progression.
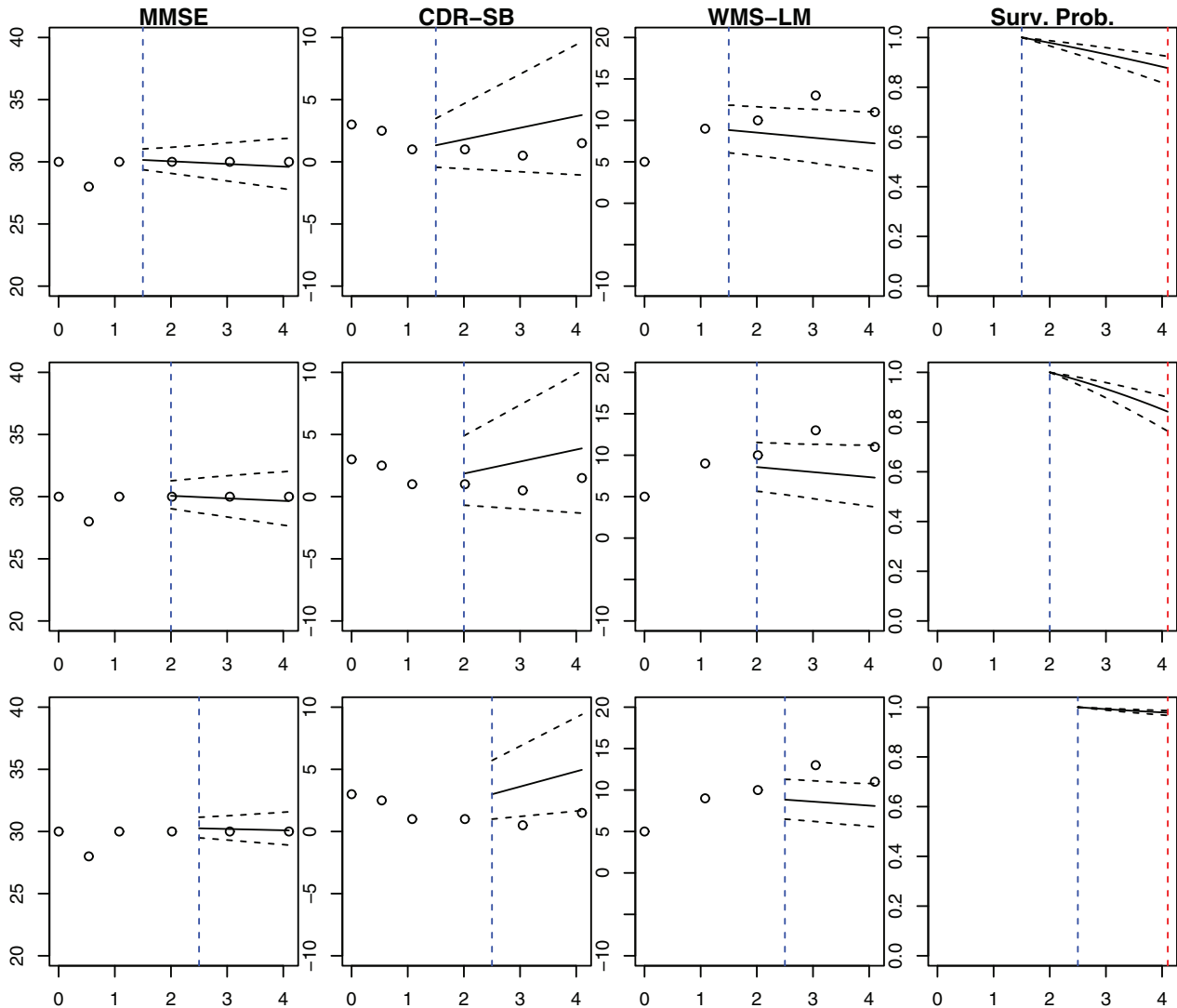
FIGURE 3   The predicted longitudinal outcome trajectories and survival probabilities for a new patient over visit time. The top, middle and bottom panels are the prediction time at 1.5, 2, 2.5 years, respectively. The circled dots are observed longitudinal outcomes. The black dotted vertical lines are the prediction time. The blue solid lines are predicted means of longitudinal outcomes and survival probabilities. The dashed curves are the corresponding 95% credible intervals. The person's censoring time is 4.10 years, which is plotted as the red-dashed vertical lines [Colour figure can be viewed at wileyonlinelibrary.com]

4.3 | Results for dynamic prediction

We conduct fourfold cross-validation and repeat 100 times to calculate the mean of estimated AUC and BS to compare the prediction accuracy for six candidate models (MJM-FD, HM, FCM, BFCM, MJM, and MJM-FCM). We set the prediction time $t = (1, 1.5, 2)$ and the prediction window $\Delta_t = (0.25, 1)$ (Figure 3).

Table 3 displays the estimated AUC and BS for models MJM-FD, HM, FCM, BFCM, MJM, and MJM-FCM. Model MJM-FD outperforms model HM for all prediction time and prediction windows with larger AUC and smaller BS, indicating using informative voxels from the whole-brain improves the predictive performance compared with the model using only hippocampus voxels. Comparing MJM-FD with FCM and BFCM, MJM-FD shows good improvement of AUC and BS in most prediction settings, suggesting the advantage of using longitudinal trajectories in the joint-model framework in AD prediction. In comparison with MJM and MJM-FCM, MJM-FD had better predictive performance at later prediction time when more longitudinal data are available, for example, larger AUC at 2 years and smaller BS at 1.5 and 2 years. Thus, the inclusion of functional predictors in both longitudinal and survival model generally improves the predictive performance of dementia diagnosis.

TABLE 3 Mean of estimated AUC and BS for MJM-FD, HM, FCM, BFCM, MJM, and MJM-FCM

| Prediction time | Δ_t | MJM-FD | | HM | | FCM | | BFCM | | MJM | | MJM-FCM | |
		AUC	BS	AUC	BS	AUC	BS	AUC	BS	AUC	BS	AUC	BS
1.0	0.25	0.797	0.055	0.763	0.054	0.745	0.054	0.808	0.052	0.801	0.053	0.795	0.050
	1.00	0.849	0.099	0.817	0.107	0.751	0.113	0.845	0.102	0.851	0.104	0.851	0.098
1.5	0.25	0.854	0.034	0.824	0.038	0.709	0.040	0.840	0.041	0.861	0.038	0.848	0.035
	1.00	0.877	0.092	0.873	0.097	0.755	0.112	0.852	0.095	0.875	0.099	0.876	0.095
2.0	0.25	0.883	0.063	0.882	0.065	0.764	0.063	0.844	0.061	0.876	0.065	0.881	0.064
	1.00	0.884	0.076	0.841	0.082	0.770	0.084	0.837	0.082	0.870	0.079	0.881	0.080

5 | SIMULATION STUDY

5.1 | Simulation setting

To evaluate the proposed MJM-FD model, we conduct a simulation study with the true parameters close to the parameter estimates in Table 2, using eight simulation settings: Setting 1 (N = 800, event rate ER = 37% with around 4.5 assessments per longitudinal outcome, missing at random (MAR) with missing rate at around 5%); Setting 2 (N = 800, ER = 37%, No Missing); Setting 3 (N = 800, ER = 63% with around 4.0 assessments per longitudinal outcome, MAR with missing rate at around 5%); Setting 4 (N = 800, ER = 63%, no missing); Setting 5-8: N = 1200, otherwise same as Setting 1-4. We generate 100 simulated datasets with a sample size of N + N/3 and set N subjects as the training set and retain the remaining subjects as the testing set. Each subject has three longitudinal outcomes (J = 3). The longitudinal outcomes are simulated at time grid $[0, \ldots, 3]$ with a 0.5 year increment. The MJM-FD model is:

$$Y_{ij}\left(t_{ijk}\right) = m_{ij}\left(t_{ijk}\right) + \epsilon_{ijk},$$

$$\text{where } m_{ij}\left(t_{ijk}\right) = \beta_{0j} + \sum_{\kappa=1}^{2}\left(t_{ijk} - \tau_\kappa\right)_+ \beta_{1j\kappa} + x_i\beta_j + \int_S g_{i,j}(s)B_j(s)ds + v_{j0}u_{i0} + v_{j1}t_{ijk}u_{i1},$$

$$h_i(t) = h_0(t)\exp\left(w_i\gamma + \int_S g_{i,w}(s)B_w(s)ds + \sum_{j=1}^{J}\alpha_j m_{ij}(t)\right),$$

where $i = 1, \ldots, N + N/3, j = 1, 2, 3, k = 1, \ldots, K_{ij}$.

For the longitudinal models, we set $(\beta_{01}, \beta_{111}, \beta_{112}, \beta_1) = (28, -1, -0.8, -0.4)$, $(\beta_{02}, \beta_{121}, \beta_{122}, \beta_2) = (2, 1, 0.8, 0.2)$, $(\beta_{03}, \beta_{131}, \beta_{132}, \beta_3) = (8, -1, -0.8, -0.2)$. The spline knot locations $\tau = (0, 2)$ and covariate $x_i \sim N(0, 1)$. The association parameters $v_2 = (-0.5, -0.4)$, $v_3 = (1, 0.5)$. We simulate $u_i \sim N(0, \Sigma)$, where $\Sigma = \begin{pmatrix} 0.25 & 0.1 \\ 0.1 & 0.25 \end{pmatrix}$. The error $\epsilon_{ijk} \sim$ Skew-Normal$(0, \omega_j^2, \delta_j)$, where $\omega = (1.2, 1, 1)$ and $\delta = (-2, 2, 0)$. We observe the noisy forms of functional predictors at domain $S = [0, \ldots, 1]$ with 0.005 increment for both longitudinal and survival outcomes. The functional predictors are simulated as: $g_{i,j}(s) = \mu_j(s) + \sum_{l=1}^{10}\phi_{jl}(s)\xi_{ijl}$, where $\mu_1(s) = 1 + s$, $\mu_2(s) = \sqrt{s+1}$, $\mu_3(s) = 1 + \exp(s)$. The eigenfunctions $\phi_{1l}(s) = \sqrt{2}\sin((2l-1)\pi s/2)$, $\phi_{2l}(s) = \sqrt{2}\cos((2l-1)\pi s/2)$, $\phi_{3l}(s) = \phi_{1l}(s)$, for $l = 1, 2, \ldots, 10$. The FPC scores $\xi_{ijl} \sim N(0, 0.9^{l-1})$ for $j = 1, 2, 3$. We observe a noisy form $g_{i,j}(s)' = g_{i,j}(s) + \epsilon_{ij}(s)$, where $\epsilon_{ij}(s) \sim N(0, 0.01)$. The coefficient functions $B_1(s) = 1 + \sin(2\pi s) + \cos(2\pi s)$, $B_2(s) = 1 + \sin(2\pi s) - \cos(2\pi s)$, $B_3(s) = 1 - \sin(2\pi s) + \cos(2\pi s)$. The missing pattern of longitudinal responses is missing at random, where missing indicator I_{ijk} is sampled from Bernoulli distribution with event probability p_{ijk}, where logit$(p_{ijk}) = 2(x_i - 2) + t_{0j}(t_{ijk} - t_{0j})$, where $t_{0j} = (2, 2.5, 1.5)$ for $j = 1, 2, 3$, generating missing proportions at around 5% for all simulation settings.

For the survival model, the baseline hazard function $h_0(t)$ is a piecewise constant function, when $0 \le t < 2$, $h_{01} = \exp(5)$, and when $t \ge 2$, $h_{02} = \exp(6)$ (when ER = 37%) or $\exp(10)$ (when ER = 63%). The covariate $w_i \sim$ Binomial$(2, 0.4)$, which mimics the APOE-ϵ4 genotype frequency, and $\gamma = 0.4$. The association parameters $\alpha_1 = -0.4$, $\alpha_2 = 0.8$, and $\alpha_3 = -0.3$. We simulate the survival probability $S_i(t)$ from standard uniform distribution $U(0, 1)$. The failure time T_i^* is computed using equations derived in the Web Supplement Section 4. The censoring time is generated from $C_i \sim U(0, 6)$.

The functional predictors are simulated as $g_{i,w}(s) = \mu_w(s) + \sum_{l=1}^{10} \phi_{wl}(s)\xi_{iwl}$, where $\mu_w(s) = 1 + \Phi(2s)$, and $\Phi(s)$ denotes the CDF of standard normal distribution. The eigenfunctions $\phi_{wl}(s) = \phi_{2l}(s)$. The FPC scores ξ_{iwl} are simulated from $N(0, 0.9^{l-1})$. We also observe a noisy form $g_{i,w}(s)' = g_{i,w}(s) + \epsilon_{iw}(s)$, where $\epsilon_{iw}(s) \sim N(0, 0.01)$. The coefficient function $B_w(s) = 1 - \sin(2\pi s) - \cos(2\pi s)$.

We perform FPCA on the observed functional predictors with the PVE criteria being 99%. We smooth the mean function using cubic B-spline function with a degree of freedom being 8. We use the tensor-product surface spline to smooth the raw covariance matrix. We perform NUTS sampling in `Stan` using two chains, each of which has a total of 2000 iterations with 1000 iterations being burn-in. The ability to estimate the true coefficient functions is assessed by the average mean squared error (AMSE), where AMSE $= 1/(100 \cdot S) \sum_{r=1}^{100} \sum_S \{B(s) - \hat{B}^{(r)}(s)\}^2$, and S is the number of grid points over domain $\boldsymbol{S}$, and $\hat{B}^{(r)}(s)$ is the estimated coefficient function at rth simulation replication. We then perform dynamic prediction with prediction time $t = (1, 1.5, 2)$ and prediction window $\Delta_t = (0.25, 0.5, 0.75, 1)$ and we calculate the true AUC, estimated AUC, true BS, and estimated BS.

5.2 | Results for simulation study

Since simulation Setting 1 is closest to the real data, we present the results from it, while the results from Setting 2-8 are presented in the Supplementary Tables 2-5. Table 4 presents bias (the average of the posterior means minus the true values), standard deviation (SD, the standard deviation of the posterior mean), standard error (SE, the square root of the average of the variance), and coverage probabilities (CP) of 95% credible intervals, in addition to AMSE for Setting 1. Table 4 suggests that the MJM-FD model performs well with small bias, SE being close to SD, the coverage probabilities being reasonably close to the nominal level of 0.95, and AMSE being close to zero. For simulation Setting 2-8, Supplementary Tables 2 and 4 also suggest parameters are estimated with small bias, and the coverage probabilities being close to 0.95. The estimated coefficient functions are close to the true coefficient functions with very little bias. Thus, the simulation studies confirm the correctness of the MJM-FD estimation method and the method in handling the missing data.

Figure 4 displays the true coefficient functions, mean of estimated coefficient functions from 100 replications, and 95% pointwise credible interval bands over domain $\boldsymbol{S}$. The results indicate that the mean of estimated coefficient functions are very close to the true coefficient functions, with 95% credible interval bands always covering the true functions.

Table 5 displays the true and estimated time-dependent AUC and BS for prediction time at $(1.5, 2, 2.5)$ years and prediction window of $(0.25, 0.5, 0.75, 1)$ years. The estimated AUC values are slightly smaller than the true AUC values, and the estimated BS are slightly larger than the true BS, indicating satisfying model performance in both discrimination and calibration. For simulation Setting 2-8, Supplementary Tables 3 and 5 suggest that the estimated AUC values are close to true AUC values, and the estimated BS are close to true BS, confirming the correctness of the dynamic prediction method under various simulation settings with different sample sizes, event rates, and with or without missing outcome data.

6 | DISCUSSION

In this article, we propose a multivariate joint model with functional data (MJM-FD) framework to jointly model multivariate longitudinal outcomes and a survival outcome in Alzheimer's disease, with the baseline magnetic resonance imaging (MRI) data being included as functional predictors. We use shared random intercepts and random slopes of time to account for subject heterogeneity at baseline and rates of progression among longitudinal outcomes. Functional principal component analysis and cubic B-spline functions are used to decompose functional predictors and coefficient functions, respectively. We adopt Bayesian methods for statistical inference and dynamic prediction and apply the MJM-FD framework to the Alzheimer's Disease Neuroimaging Initiative (ADNI) study, where missing longitudinal responses are imputed in the Markov-Chain Monte Carlo (MCMC) sampling procedure. For longitudinal outcomes, we conclude that older ages are significantly associated with worse memory functions (lower Wechsler Memory Scale—logical memory scores, WMS-LM), and lower education is significantly associated with worse cognition and memory functions (lower Mini-Mental State Examination and WMS-LM scores). Higher Clinical Dementia Rating—sum of boxes scores and lower WMS-LM scores are significantly associated with higher Alzheimer's disease (AD) risk among participants with mild cognitive impairment (MCI). The atrophy of voxels inside hippocampus and fornix column and body region is associated with worse cognition scores and memory function. The atrophy of voxels inside hippocampus region and other ROIs may expedite the progression from MCI to AD. We also construct a

TABLE 4 Parameter estimation from 100 simulation replications for simulation Setting 1

	Parameters	Bias	SD	SE	CP
Longitudinal outcomes					
Y_1	$\beta_{01} = 28$	−0.013	0.056	0.059	0.950
	$\beta_{111} = -1$	−0.003	0.029	0.030	0.960
	$\beta_{112} = -0.8$	0.003	0.080	0.086	0.950
	$\beta_1 = -0.4$	−0.001	0.024	0.026	0.980
Y_2	$\beta_{02} = 2$	−0.001	0.037	0.038	0.980
	$\beta_{121} = 1$	0.002	0.018	0.019	0.940
	$\beta_{122} = 0.8$	−0.001	0.061	0.065	0.970
	$\beta_2 = 0.2$	−0.001	0.015	0.016	0.940
Y_3	$\beta_{03} = 8$	−0.060	0.217	0.318	0.980
	$\beta_{131} = -1$	0.001	0.032	0.029	0.910
	$\beta_{132} = -0.8$	−0.011	0.113	0.099	0.910
	$\beta_3 = -0.2$	0.000	0.021	0.026	0.980
	$\sigma_{u0} = 0.5$	−0.000	0.028	0.028	0.960
	$\sigma_{u1} = 0.5$	−0.002	0.025	0.024	0.950
	$\rho = 0.4$	−0.000	0.075	0.072	0.940
	$v_{20} = -0.5$	−0.008	0.040	0.039	0.930
	$v_{21} = -0.4$	0.003	0.023	0.025	0.960
	$v_{30} = 1$	0.009	0.067	0.067	0.950
	$v_{31} = 0.5$	−0.003	0.039	0.039	0.970
	$\delta_1 = -2$	0.026	0.167	0.177	0.960
	$\delta_2 = 2$	0.024	0.144	0.149	0.930
	$\delta_3 = 0$	0.035	0.289	0.419	0.980
Survival outcomes					
	$\gamma = 0.4$	−0.002	0.096	0.095	0.960
	$\alpha_1 = -0.4$	−0.014	0.046	0.044	0.940
	$\alpha_2 = 0.8$	0.023	0.063	0.058	0.910
	$\alpha_3 = -0.3$	−0.000	0.051	0.046	0.910
	AMSE.B1	0.019			
	AMSE.B2	0.008			
	AMSE.B3	0.016			
	AMSE.BW	0.184			

dynamic prediction framework to predict future longitudinal outcomes and risk of AD onset for new subjects. The inclusion of selected voxels from the whole voxel-based baseline MRI data and the inclusion of longitudinal trajectories in the survival model improve the overall model fitting and predictive performance, comparing with other candidate models.

There are few limitations we may address as future directions. First, we only use the baseline MRI data as functional predictors in the MJM-FD, while the longitudinal MRI data are available for some individuals. The longitudinal MRI data can be treated as two-dimensional functional predictors with time and volumes of voxels being two-dimensional domains and the coefficient functions can be constructed in the two-dimensional domains. Second, although we utilize
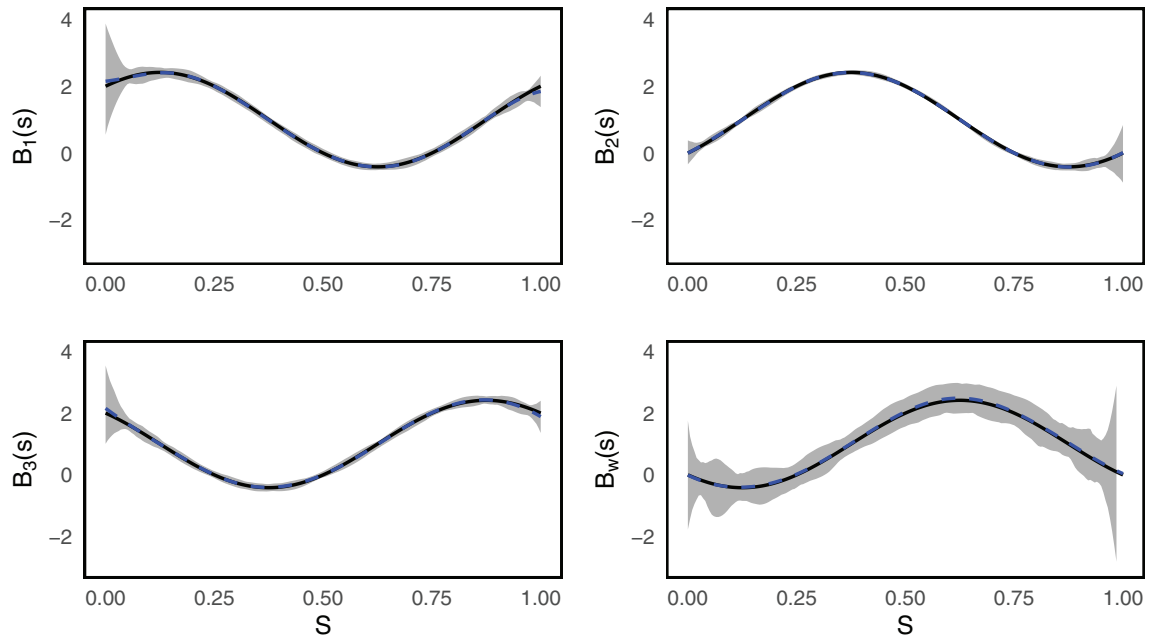
FIGURE 4 The true coefficient functions (solid curves), mean of estimated coefficient functions (dashed curves), and 95% pointwise credible interval bands over domain $\boldsymbol{S}$. The top left, top right, bottom left, and bottom right panels are the coefficient functions $B_1(s)$, $B_2(s)$, $B_3(s)$, $B_w(s)$, respectively [Colour figure can be viewed at wileyonlinelibrary.com]

TABLE 5 True area under the ROC curve (AUC), estimated AUC, true Brier score (BS), and estimated BS from 100 simulation replications

t	Δ_t	True AUC	Estimated AUC	True BS	Estimated BS
1.0	0.25	0.895	0.885	0.022	0.022
	0.50	0.893	0.880	0.041	0.044
	0.75	0.893	0.879	0.059	0.063
	1.00	0.894	0.879	0.074	0.080
1.5	0.25	0.872	0.864	0.032	0.033
	0.50	0.877	0.868	0.056	0.059
	0.75	0.890	0.877	0.087	0.093
	1.00	0.900	0.884	0.106	0.117
2.0	0.25	0.878	0.867	0.064	0.067
	0.50	0.886	0.875	0.100	0.107
	0.75	0.902	0.887	0.114	0.124
	1.00	0.919	0.901	0.064	0.071

the whole-brain MRI data, we only select voxels with FDR-adjusted p-values less than 0.05 for MMSE and CDR-SB, and 0.01 for WMS-LM and survival outcomes, which is an ad hoc method for voxel selection. A recent article discusses a soft-thresholded Gaussian process for scalar-on-image regression and variable selection and it is an alternative way for voxel selection.[58] In addition, the informative voxels in WMS-LM are 1387, about 20 fold more than the informative voxels in CDR-SB (69 voxels), indicating that WMS-LM has larger voxel volume variation than CDR-SB. Finally, we only use one of the representative genetic variants of AD progression, APOE-ϵ4. The MJM-FD framework can be extended to account for multiple genetic variants by adopting the polygenic risk scores (PRS) approach.[59,60] Furthermore, age is found to be insignificantly associated with AD progression. The possible reason is that age is a shared covariate in the longitudinal and

survival model, and there exists a direct effect and an indirect effect for age. Thus, the age effect is possibly ameliorated in the survival model.

As pointed out by one reviewer, our MJM-FD assumes that the effects of baseline brain structure on the longitudinal trajectories and survival outcome are time-independent. To relax this assumption, we may incorporate coefficient functions with two dimensions in voxel space and time, for example, replacing $B_j(s)$ in Model (2) by $B_j(s, t_{ijk})$. This modification allows the baseline brain structure to be associated with the intensity of the cognitive change. On the other hand, this change greatly increases the model complexity and computational burden, and we would like to investigate it in our future research. Furthermore, in the longitudinal submodel of the MJM-FD, we model the correlation across multivariate longitudinal outcomes using the shared random effects u_{i0} and u_{i1}. A more general form is to include outcome-specific random effects which follow a multivariate normal distribution. The large number of required random effects may impose challenging numerical difficulty and computational burden.[61] A practical solution to reduce computational burden is to adopt a two-stage approach and use a correction factor based on the importance sampling theory. The correction factor allows us to reweight each realization of the MCMC sample such that the estimates more closely approximate those obtained via the full multivariate joint model.[7]

In this article, we use the scalar-on-function regression method to include functional predictors in the multivariate joint model framework. It would be more general if the longitudinal outcome data are treated as functional outcomes on a sparse time domain. To this end, we may treat longitudinal observations as functional outcomes and MRI data as functional predictors, and adopt the function-on-function regression methods. Specifically, the functional mixed effects models can be readily applied to model the longitudinal outcomes and the latent mean function can be incorporated in the survival model as described in several articles.[62-64] The dynamic prediction framework can be also developed.

ORCID

Haotian Zou https://orcid.org/0000-0002-3595-8716
Kan Li https://orcid.org/0000-0002-5762-7440
Donglin Zeng https://orcid.org/0000-0003-0843-9280
Sheng Luo https://orcid.org/0000-0003-4214-5809

REFERENCES

1. Hebert LE, Weuve J, Scherr PA, Evans DA. Alzheimer disease in the United States (2010–2050) estimated using the 2010 census. *Neurology*. 2013;80:1778-1783.
2. Jack CR, Albert MS, Knopman DS, et al. Introduction to the recommendations from the national institute on aging-Alzheimer's association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimers Dement*. 2011;7:257-262.
3. Gross AL, Mungas DM, Leoutsakos JS, Albert MS, Jones RN. Alzheimer's disease severity, objectively determined and measured. *Alzheimers Dement*. 2016;4:159-168.
4. Mitchell AJ, Shiri-Feshki M. Rate of progression of mild cognitive impairment to dementia: meta-analysis of 41 robust inception cohort studies. *Acta Psychiatr Scand*. 2009;119:252-265.
5. Wulfsohn MS, Tsiatis AA. A joint model for survival and longitudinal data measured with error. *Biometrics*. 1997;53:330-339.
6. Henderson R, Diggle P, Dobson A. Joint modelling of longitudinal measurements and event time data. *Biostatistics*. 2000;1:465-480.
7. Mauff K, Steyerberg E, Kardys I, Eric B, Dimitris R. Joint models with multiple longitudinal outcomes and a time-to-event outcome: a corrected two-stage approach. *Stat Comput*. 2020;30:999-1014.
8. Long JD, Mills JA. Joint modeling of multivariate longitudinal data and survival data in several observational studies of Huntington's disease. *BMC Med Res Methodol*. 2018;18:138.
9. Proust-Lima C, Dartigues JF, Jacqmin-Gadda H. Joint modeling of repeated multivariate cognitive measures and competing risks of dementia and death: a latent process and latent class approach. *Stat Med*. 2016;35:382-398.

10. Hickey GL, Philipson P, Jorgensen A, Kolamunnage-Dona R. Joint modelling of time-to-event and multivariate longitudinal outcomes: recent developments and issues. *BMC Med Res Methodol*. 2016;16:1-15.

11. Papageorgiou G, Mauff K, Tomer A, Rizopoulos D. An overview of joint modeling of time-to-event and longitudinal outcomes. *Annu Rev Stat Appl*. 2019;6:223-240.

12. Rizopoulos D. Dynamic predictions and prospective accuracy in joint models for longitudinal and time-to-event data. *Biometrics*. 2011;67:819-829.

13. Li L, Luo S, Hu B, Greene T. Dynamic prediction of renal failure using longitudinal biomarkers in a cohort study of chronic kidney disease. *Stat Biosci*. 2017;9:357-378.

14. Schumacher M, Hieke S, Ihorst G, Monika E. Dynamic prediction: a challenge for biostatisticians, but greatly needed by patients, physicians and the public. *Biomed J*. 2020;62:822-835.

15. Leroux A, Xiao L, Crainiceanu CM, William C. Dynamic prediction in functional concurrent regression with an application to child growth. *Stat Med*. 2018;37:1376-1388.

16. Vemuri P, Jack CR. Role of structural MRI in Alzheimer's disease Alzheimers. *Res Ther*. 2010;2:23.

17. Lerch JP, Kouwe AJ, Raznahan A, et al. Studying neuroanatomy using MRI. *Nat Neurosci*. 2017;20:314-326.

18. Nicole RG, Vince DC, Godfrey DP, Francis A, Buchanan RW. Voxel-based morphometry versus region of interest: a comparison of two methods for analyzing gray matter differences in schizophrenia. *Schizophr Res*. 2005;74:135-147.

19. Buchanan RW, Francis A, Arango C, et al. Morphometric assessment of the heteromodal association cortex in schizophrenia. *Am J Psychiatry*. 2004;161:322-331.

20. Nho K, Risacher SL, Crane PK, et al. Voxel and surface-based topography of memory and executive deficits in mild cognitive impairment and Alzheimer's disease. *Brain Imaging Behav*. 2012;6:551-567.

21. Jennifer LW. Voxel-based morphometry: an automated technique for assessing structural changes in the brain. *J Neurosci*. 2009;29:9661-9664.

22. Wang J, Chiou J, Müller H. Functional data analysis. *Annu Rev Stat Appl*. 2016;3:257-295.

23. Morris JS. Functional regression. *Annu Rev Stat Appl*. 2015;2:321-359.

24. Goldsmith J, Bobb J, Crainiceanu CM, Caffo B, Reich D. Penalized functional regression. *J Comput Graph Stat*. 2010;20:830-851.

25. Goldsmith J, Crainiceanu CM, Caffo B, Reich D. Longitudinal penalized functional regression for cognitive outcomes on neuronal tract measurements. *J R Stat Soc Ser C Appl Stat*. 2012;61:453-469.

26. Reiss PT, Goldsmith J, Shang HL, Ogden RT. Methods for scalar-on-function regression. *Int Stat Rev*. 2017;85:228-249.

27. Yao F, Müller H, Wang J. Functional data analysis for sparse longitudinal data. *J Am Stat Assoc*. 2005;100:577-590.

28. Happ C, Greven S. Multivariate functional principal component analysis for data observed on different (dimensional) domains. *J Am Stat Assoc*. 2018;113:649-659.

29. Di CZ, Crainiceanu CM, Caffo BS, Punjabi NM. Multilevel functional principal component analysis. *Ann Appl Stat*. 2009;3:458-488.

30. Xiao L, Zipunnikov V, Ruppert D, Crainiceanu CM. Fast covariance estimation for high-dimensional functional data. *Stat Comput*. 2016;26:409-421.

31. Li K, Luo S. Dynamic predictions in Bayesian functional joint models for longitudinal and time-to-event data: an application to Alzheimer's disease. *Stat Methods Med Res*. 2019;28:327-342.

32. Li K, Chan W, Doody RS, Rachelle SD, Joseph Q, Sheng L. Prediction of conversion to Alzheimer's disease with longitudinal measures and time-to-event data. *J Alzheimers Dis*. 2017;58:360-370.

33. Besser L, Kukull W, Knopman DS, et al. Version 3 of the national Alzheimer's coordinating center's uniform data set. *Alzheimer Dis Assoc Disord*. 2018;32:351-358.

34. Bennett DA, Buchman AS, Boyle PA, Barnes LL, Wilson RS, Schneider JA. Religious orders study and rush memory and aging project. *J Alzheimers Dis*. 2018;64:S161-S189.

35. Ellis KA, Bush AI, Darby D, et al. The Australian imaging, biomarkers and lifestyle (AIBL) study of aging: methodology and baseline characteristics of 1112 individuals recruited for a longitudinal study of Alzheimer's disease. *Int Psychogeriatr*. 2009;21:672-687.

36. Leoutsakos JM, Forrester SN, Corcoran CD, et al. Latent classes of course in Alzheimer's disease and predictors: the cache county dementia progression study. *Int J Geriatr Psychiatry*. 2015;30:824-832.

37. Chapman KR, Bing-Canar H, Alosco ML, et al. Mini mental state examination and logical memory scores for entry into Alzheimer's disease trials Alzheimers. *Res Ther*. 2016;8:1-11.

38. Ashburner J, Friston KJ. Voxel-based morphometry-the methods. *Neuroimage*. 2000;11:805-821.

39. Pergher V, Demaerel P, Soenen O, et al. Identifying brain changes related to cognitive aging using VBM and visual rating scales. *Neuroimage Clin*. 2019;22:101697.

40. Oschwald J, Guye S, Liem F, et al. Brain structure and cognitive ability in healthy aging: a review on longitudinal correlated change. *Rev Neurosci*. 2020;31:1-57.

41. Anatürk M, Suri S, Zsoldos E, et al. Associations between longitudinal trajectories of cognitive and social activities and brain health in old age. *JAMA Netw Open*. 2020;3:e2013793.

42. Chou YY, Leporé N, Saharan P, et al. Ventricular maps in 804 ADNI subjects: correlations with CSF biomarkers and clinical decline. *Neurobiol Aging*. 2010;31:1386-1400.

43. Evans MC, Barnes J, Nielsen C, et al. Volume changes in Alzheimer's disease and mild cognitive impairment: cognitive associations. *Eur Radiol*. 2010;20:674-682.

44. Kovacevic S, Rafii MS, Brewer JB. High-throughput, fully-automated volumetry for prediction of MMSE and CDR decline in mild cognitive impairment Alzheimer. *Dis Assoc Disord*. 2009;23:139.

45. Armstrong NM, An Y, Shin JJ, et al. Associations between cognitive and brain volume changes in cognitively normal older adults. *Neuroimage*. 2020;223:117289.

46. Geert V, Steffen F, Geert M, Davidian M. The analysis of multivariate longitudinal data: a review. *Stat Methods Med Res*. 2014;23:42-59.

47. Lawless JF, Zhan M. Analysis of interval-grouped recurrent-event data using piecewise constant rate functions. *Can J Stat*. 1998;26:549-565.

48. Feng S, Wolfe RA, Port FK. Frailty survival model analysis of the national deceased donor kidney transplant dataset using Poisson variance structures. *J Am Stat Assoc*. 2005;100:728-735.

49. Liu L, Huang X. Joint analysis of correlated repeated measures and recurrent events processes in the presence of death, with application to a study on acquired immune deficiency syndrome. *J R Stat Soc Ser C Appl Stat*. 2009;58:65-81.

50. Homan MD, Gelman A. The No-U-Turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo. *J Mach Learn Res*. 2014;15:1593-1623.

51. Vehtari A, Gelman A, Gabry J. Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Stat Comput*. 2017;27:1413-1432.

52. Li L, Greene T, Hu B. A simple method to estimate the time-dependent receiver operating characteristic curve and the area under the curve with right censored data. *Stat Methods Med Res*. 2018;27:2264-2278.

53. Graf E, Schmoor C, Sauerbrei W, Martin S. Assessment and comparison of prognostic classification schemes for survival data. *Stat Med*. 1999;18:2529-2545.

54. Oishi K, Faria A, Mori S. JHU-MNI-ss Atlas; 2010.

55. Oishi K, Faria A, Jiang H, et al. Atlas-based whole brain white matter analysis using large deformation diffeomorphic metric mapping: application to normal elderly and Alzheimer's disease participants. *Neuroimage*. 2009;46:486-499.

56. Jack CR, Petersen RC, Xu YC, et al. Prediction of AD with MRI-based hippocampal volume in mild cognitive impairment. *Neurology*. 1999;52:1397-1397.

57. Xia M, Wang J, He Y. BrainNet viewer: a network visualization tool for human brain connectomics. *PLoS One*. 2013;8:e68910.

58. Kang J, Reich BJ, Staicu A. Scalar-on-image regression via the soft-thresholded Gaussian process. *Biometrika*. 2018;105:165-184.

59. Escott-Price V, Sims R, Bannister C, et al. Common polygenic variation enhances risk prediction for Alzheimer's disease. *Brain*. 2015;138:3673-3684.

60. Chaudhury S, Brookes KJ, Patel T, et al. Alzheimer's disease polygenic risk score as a predictor of conversion from mild-cognitive impairment. *Transl Psychiatry*. 2019;9:154.

61. Vasdekis GS, Rizopoulos D, Moustaki I. Weighted pairwise likelihood estimation for a general class of random effects models. *Biostatistics*. 2014;15:677-689.

62. Guo W. Functional mixed effects models. *Biometrics*. 2002;58:121-128.

63. Guo W. Functional mixed effects models. *Stat Methods Med Res*. 2004;13:49-62.

64. Yao F. Functional principal component analysis for longitudinal and survival data. *Stat Sin*. 2007;17:965-983.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.