Received: 25 November 2020 | Revised: 15 September 2021 | Accepted: 15 September 2021

DOI: 10.1002/sim.9241

RESEARCH ARTICLE

Statistics in Medicine WILEY

A joint model for multivariate longitudinal and survival data to discover the conversion to Alzheimer's disease

Kai Kang[1] | **Deng Pan**[2] | **Xinyuan Song**[1]

[1]Department of Statistics, The Chinese University of Hong Kong, Hong Kong, China

[2]School of Mathematics and Statistics, Huazhong University of Science and Technology, Wuhan, China

Correspondence
Deng Pan, School of Mathematics and Statistics, Huazhong University of Science and Technology, Wuhan, China.
Email: pand@hust.edu.cn
Xinyuan Song, Department of Statistics, The Chinese University of Hong Kong, Hong Kong, China.
Email: xysong@sta.cuhk.edu.hk

Abstract

Alzheimer's disease (AD) is an incurable and progressive disease that starts from mild cognitive impairment and deteriorates over time. Examining the effects of patients' longitudinal cognitive decline on time to conversion to AD and obtaining a reliable diagnostic model are therefore critical to the evaluation of AD prognosis and early treatment. Previous studies either assess patients' cognitive impairment through a single cognitive test or assume it changes linearly across time, thereby leading to an incomplete measure of cognitive decline or overlooking the subtle trajectory pattern of patients' cognitive impairment. This study develops a new joint model to address these shortcomings. First, a dynamic factor analysis model is adopted to characterize cognitive impairment through multiple cognitive measures in a comprehensive manner. Second, a spline-based random coefficient model is proposed to reveal possibly nonlinear trajectories of patients' cognitive decline. Finally, a proportional hazard model is considered to examine the effects of time-invariant markers and time-variant cognitive impairment on AD hazards. A Bayesian approach coupled with spline approximation techniques and MCMC methods is developed to conduct statistical inference. The application of the proposed method to the Alzheimer's Disease Neuroimaging Initiative study provides new insights into the prevention of AD and shows a high prediction capacity of the proposed method.

KEYWORDS

factor analysis, latent trajectory model, longitudinal responses, MCMC methods, time-to-event outcome

1 | INTRODUCTION

Alzheimer's disease (AD) is a chronic and progressive disease that usually starts with a short memory loss, also referred to as mild cognitive impairment (MCI), and worsens over time. Clinical trials or prospective studies investigating AD often recruit individuals with MCI and repeatedly collects measurements of cognitive impairment via neuropsychological or behavioral assessments over time. During the cohort study, patients who meet the specific inclusion criteria are diagnosed as having AD, and thus, a possibly censored time-to-AD outcome is also recorded. Only a portion of MCI patients progress to AD, whereas some individuals remain stable or even revert to normal cognition. Thus, exploring the

association between the feature of longitudinal cognitive impairment and time-to-AD and investigating early markers for the diagnosis of AD are of great value in AD prevention and targeted treatment.

The Alzheimer's Disease Neuroimaging Initiative (ADNI) study began at 2004, and it tracked the pathology of AD in the human brain with serial magnetic resonance imaging (MRI), biospecimen biomarkers, and various clinical and neurocognitive measures from subjects with normal aging, MCI, dementia, or AD. Detailed information about ADNI, including complete study visit procedures and inclusion/exclusion criteria, can be found in the official website (www.adni-info.org). After obtaining informed consent, participants underwent a series of initial tests, including Alzheimer Disease Assessment Scale-Cognitive 11 (ADAS11), Alzheimer Disease Assessment Scale-Cognitive 13 (ADAS13), and Functional Assessment Questionnaire (FAQ), at baseline and were reassessed through interviews over subsequent years. Although these tests all assess cognitive impairment, their main focuses vary. ADAS11 consists of 11 items that correspond to words, spoken language, and simple commands, ADAS13 incorporates delayed recall and digit cancellation tasks further, and FAQ mainly examines the ability to complete daily life activities, such as preparing a balanced meal or traveling outside one's neighborhood. During the multiple study phases that were conducted for 9 years, some patients were quickly diagnosed as having AD or other dementia-related diseases based on the specific inclusion criteria, whereas some patients reflected a normal aging pattern or even remained stable. In this longitudinal study, our primary goal is to examine the effects of time-invariant markers and time-variant cognitive impairment on AD hazards. Toward this goal, the development of a statistical model that summarizes the longitudinal cognitive decline and it's association with the progression of AD is required.

Joint modeling of longitudinal and survival data[1,2] is a popular framework for investigating the relationship between repeated measurements and time-to-event outcomes. Such joint models basically consist of two parts, namely, a mixed-effects model for characterizing the trajectory of longitudinal measures and a survival model, such as a proportional hazard (PH) model,[3] to link the underlying random effects to the survival of interest. Owing to the superiority to simultaneously reveal the structure of repeated measurements and it's association with time-to-event process, joint modeling of longitudinal and survival data has attracted significant attention in medical research.[4-7] Such kinds of joint models have previously been applied to the analysis of repeated measurement of cognitive decline and time-to-AD. For instance, Hashemi et al[8] used a joint modeling approach to analyze dementia and the score of a psychometric test, namely, the Mini Mental State Examination (MMSE). Jacqmin-Gadda et al[9] developed a random change-point model to jointly model the score of Benton Visual Retention Test and dementia. Recently, Li and Luo[10] incorporated functional data into joint modeling and demonstrated the association between time-to-AD and longitudinal measurement of ADAS11.

In the ADNI study, cognitive impairment is characterized by multiple cognitive measures from various perspectives. However, existing methods assess patients' cognitive impairment through a single cognitive test. The analysis in Section 4 shows that such an incomplete measurement of cognitive impairment reduces model interpretability and prediction accuracy. Moreover, previous studies (eg, Ibrahim et al[5]) restrict individual trajectories to be a linear form with random intercept and time slope. Such a linearity assumption is violated in the ADNI study and shown in Section 4 to result in lower discriminative capacity in the diagnosis of AD. We propose a new joint model to address these shortcomings. First, we introduce a dynamic factor analysis model to characterize time-variant cognitive impairment using ADAS11, ADAS13, and FAQ together. The introduction of factor analysis can integrate the information reflected by various cognitive assessments, but avoids the multicollinearity problem caused by their high correlations, thereby eliminating information loss, reducing measurement errors, and enhancing model interpretability. Then, we establish a spline-based random coefficient trajectory model to reveal possibly nonlinear trajectories of patients' cognitive decline. Such a nonparametric model relaxes the linearity assumption in the previous studies and is thus highly flexible in capturing the complex patterns of cognitive impairment trajectories. Finally, we incorporate the trajectory function into a PH model to examine the effects of time-invariant markers and time-variant cognitive impairment on AD hazards. We develop a Bayesian approach coupled with an efficient Markov chain Monte Carlo (MCMC) algorithm to perform statistical inference because it shows powerful and efficient management of complex models with latent variables and allows for incorporating useful prior knowledge.

In Section 2, we introduce the proposed joint model with multivariate longitudinal and survival data. Associated model identifiability issues are also discussed. Section 3 presents the Bayesian approach. Prior specification and posterior inference are also described. Section 4 applies the proposed method to the ADNI dataset. Section 5 demonstrates the empirical performance of the proposed method through simulation studies. Section 6 concludes the study. Technical details are provided in the Web Appendix.

2 | JOINT MODELS

We propose a joint model that involves three components. The first component is a dynamic confirmatory factor analysis (CFA) model, the second component is a spline-based random coefficient trajectory model, and the third component is a PH model.

For subject i and jth measurement, let $\mathbf{y}_{ij}(t)$ be a $p \times 1$ vector of the observed variables at time t. For brevity, we denote $\mathbf{y}_{ij} = \mathbf{y}_{ij}(t_{ij})$ for $i = 1, \ldots, n$ and $j = 1, \ldots, m_i$, where m_i is the number of measurements for subject i. The dynamic CFA model used to characterize time-variant latent variables through multivariate time-variant observed variables is defined as follows:

$$\mathbf{y}_{ij} = \mathbf{B}\boldsymbol{\eta}_{ij} + \boldsymbol{\epsilon}_{ij}, \quad j = 1, \ldots, m_i, \tag{1}$$

where $\mathbf{B}$ is a $p \times q$ ($p > q$) matrix of factor loadings; $\boldsymbol{\eta}_{ij}$ is a $q \times 1$ vector of latent variables for subject i at time j; and $\boldsymbol{\epsilon}_{ij}$ is a $p \times 1$ vector of random errors independent of $\boldsymbol{\eta}_{ij}$ and distributed as $N(\mathbf{0}, \boldsymbol{\Psi})$ with diagonal matrix $\boldsymbol{\Psi}$. In substantive research, we may obtain prior knowledge about p, q, and $\mathbf{B}$ based on the objectives of the study, the meaning of the observed variables suggested by subject experts, and/or the existing literature. For example, in the ADNI study in Section 5, the existing medical literature suggested that cognitive impairment is measured using several neuropsychological or behavioral assessments, including ADAS11, ADAS13, and FAQ. Thus, $p = 3$, $q = 1$, and $\mathbf{B}$ is a 3×1 factor loading matrix.

Henceforth, we assume a unidimensional η_{ij} ($q = 1$) to simplify notations. An extension to the case of $q > 1$ is straightforward. For the longitudinal latent variable η_{ij}, we consider a random coefficient model with splines, as follows:

$$\eta_{ij} = h_i(t_{ij}) = \sum_{l=1}^{L} u_{il} W_l(t_{ij}), \tag{2}$$

where $\mathbf{u}_i = (u_{i1}, u_{i2}, \ldots, u_{iL})^T$ is a $L \times 1$ vector of time-invariant subject-specific random coefficients; $\mathbf{u}_i \sim N(\mathbf{0}, \boldsymbol{\Phi})$, $\boldsymbol{\Phi}$ is an unknown covariance matrix; $W_l(\cdot)s$ are basis functions, such as piecewise polynomials or natural cubic splines,[11] and L is the number of basis functions that are used to estimate the smoothing individual trajectory $h_i(\cdot)$. Previous studies[12,13] indicate that a small number of knots (eg, $L \in [5, 10]$) are sufficient to provide a good approximation for functions with moderate curvature. It is common to choose L to be less than or equal to the maximum number of measurements because we only have observations in these discrete follow-up time points. Alternatively, one can also set a relatively large L and then use penalized splines[14,15] or adopt model-selection criteria (eg, BIC or DIC) to choose the number of knots to avoid overfitting. Such a personalized model framework enables sufficient flexibility to specify the potential nonlinear or fluctuated longitudinal trajectories, in particular, longitudinal trajectories of cognitive impairment.

We follow the existing literature of ADNI study[10,16,17] to consider a PH model to investigate the potential risk factors of AD hazards. For subject i, let T_i denote the failure time of interest, $\mathbf{z}_i$ be an $r \times 1$ vector of the observed covariates at the baseline, and C_i be the censoring time that is conditionally independent of T_i given $\mathbf{z}_i$ and $\mathbf{u}_i$. Instead of observing all T_i for all i, we observe $V_i = \min(T_i, C_i)$. The failure indicator is denoted as $\Delta_i = I(T_i \leq C_i)$, where $I(\cdot)$ is the indicator function. The PH model is defined as follows:

$$\lambda(t|\mathbf{z}_i, \mathbf{u}_i) = \lambda_0(t) \exp\left(\boldsymbol{\beta}^T \mathbf{z}_i + \alpha h_i(t)\right), \tag{3}$$

where $\lambda(t|\cdot)$ is the hazard function of T_i, $\lambda_0(t)$ is an unspecified baseline hazard function, $\boldsymbol{\beta}$ is a $r \times 1$ vector of the regression parameters, and α is a scale parameter to adjust the effect of individual trajectories of latent variables on the hazard function. The PH model (3) enables the assessment of time-invariant and time-variant risk factors. For instance, in the ADNI example in Section 4, the effects of baseline covariates, including marital status, gender, education level, and APOE-ϵ4 genotypes, and time-variant cognitive impairment on the risk of AD can be simultaneously examined though the proposed PH model.

The proposed model includes variables $(\mathbf{y}_{ij}, \mathbf{z}_i, V_i, \Delta_i, \eta_{ij}, \mathbf{u}_i)$, in which $(\mathbf{y}_{ij}, \mathbf{z}_i, V_i, \Delta_i)$ are observed and $(\eta_{ij}, \mathbf{u}_i)$ are latent. While CFA model (1) measures η_{ij} based on the information of $\mathbf{y}_{ij}$, trajectory model (2) constructs varying-coefficient splines with random effect $\mathbf{u}_i$ to describe the trajectory function of η_{ij}, $h_i(t)$, and PH model (3) examines the effects of $h_i(t)$ and baseline covariate $\mathbf{z}_i$ on survival outcome (V_i, Δ_i). Compared with results of the existing studies on the

progression to AD, the proposed joint model defined by (1)-(3) can accommodate the following additional features. First, the dynamic CFA model (1) measures latent traits through multiple observed variables over time. Such factor analysis technique provides a feasible methodology for a comprehensive characterization and attractive interpretation of latent cognitive impairment, eliminates possible multicollinearity induced by highly correlated observed cognitive assessments, and reduces model dimensionality without losing information.[18-22] Second, the latent trajectory model (2) determines flexibly the patterns of individual trajectories and addresses within-subject dependency through subject-specific random coefficients. The spline-based specification provides sufficient flexibility to mostly uncover individual trajectories of cognitive decline during the long-term cohort study. Lastly, the PH model (3) not only examines how time-invariant covariates influence the risk of AD but also assesses the dynamic effect of individual trajectories of cognitive decline on time to conversion to AD.

2.1 | Model identifiability

The dynamic CFA model (1) is unidentifiable without imposing appropriate identifiability constraints on the parameters. Specifically, for an arbitrary non-singular matrix $\mathbf{M}$, we have

$$\mathbf{y}_{ij} = \mathbf{B}\boldsymbol{\eta}_{ij} + \boldsymbol{\epsilon}_{ij} = \mathbf{B}\mathbf{M}\mathbf{M}^{-1}\boldsymbol{\eta}_{ij} + \boldsymbol{\epsilon}_{ij} = \mathbf{B}^{*}\boldsymbol{\eta}^{*}_{ij} + \boldsymbol{\epsilon}_{ij}, \tag{4}$$

where $\mathbf{B}^{*} = \mathbf{B}\mathbf{M}$, $\boldsymbol{\eta}^{*}_{ij} = \mathbf{M}^{-1}\boldsymbol{\eta}_{ij} \sim N[\mathbf{0}, \mathbf{M}^{-1}\mathbf{W}^{T}_{ij}\boldsymbol{\Phi}\mathbf{W}_{ij}(\mathbf{M}^{-1})^{T}]$, and $\mathbf{W}_{ij} = (W_1(t_{ij}), \ldots, W_L(t_{ij}))^{T}$, indicating that parameters $\mathbf{B}$ and $\boldsymbol{\Phi}$ are not simultaneously estimable without imposing identifiability constraint. In the current study, we follow a common practice in latent variable models[23] to fix additional elements in $\mathbf{B}$ at preassigned values, such that the only possible matrix satisfying (4) is the identity matrix. Details of such identifiability constrains and an illustrated example are discussed in the Web Appendix.

3 | BAYESIAN ANALYSIS

3.1 | Estimation

The baseline hazard function $\lambda_0(t)$ is unknown, and providing the evidence of a parametric hazard function shape is usually difficult in substantive studies. We follow the existing literature (eg, Ibrahim et al,[24] Yin and Ibrahim,[25] Lee et al,[16] Pan et al[26]) to approximate $\lambda_0(t)$ through a commonly used piecewise constant distribution. One can also use B-splines[27,28] to obtain a more flexible and smooth estimate of $\lambda_0(t)$ (eg, Zhou et al[29]). In this study, we assume that $\lambda_0(t)$ is piecewise constant, as follows:

$$\lambda_0(t) = \lambda_g, \quad \text{for} \quad s_{g-1} < t \leq s_g, \quad g = 1, \ldots, G, \tag{5}$$

where $0 = s_0 < s_1 < \cdots < s_G$ define the intervals for $\lambda_0(t)$ and are selected according to the quantiles of T_i with $s_G > \max_i T_i$ for $i = 1, \ldots, n$; and $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_G)^{T}$. We define $v_{ig} = 1$ if $T_i \in (s_{g-1}, s_g]$ (ie, subject i fails or is censored in the gth interval) and 0 otherwise.

Denote $\boldsymbol{\theta} = (\alpha, \boldsymbol{\beta}, \boldsymbol{\lambda}, \mathbf{B}, \boldsymbol{\Phi}, \boldsymbol{\Psi})$. We first specify the prior distribution for each element of $\boldsymbol{\theta}$. Similar to previous studies,[21,25] we consider the prior distributions for parameters $\boldsymbol{\lambda}$, $\boldsymbol{\beta}$, and α as follows. For $g = 1, \ldots, G$,

$$\lambda_g \sim Gamma(a_1, a_2), \quad \boldsymbol{\beta} \sim N(\boldsymbol{\mu}_{\beta}, \sigma^2_{\beta}\mathbf{I}_r), \quad \alpha \sim N(\mu_{\alpha}, \sigma^2_{\alpha}), \tag{6}$$

where $a_1, a_2, \boldsymbol{\mu}_{\beta}, \sigma^2_{\beta}, \mu_{\alpha}$, and σ^2_{α} are hyperparameters with pre-assigned values, and $\mathbf{I}_r$ is the r-dimensional identity matrix. Alternatively, we can also assign prior correlation among λ_gs using a prior $\boldsymbol{\varphi} \sim N(\boldsymbol{\mu}_{\phi}, \boldsymbol{\Sigma}_{\phi})$, where $\varphi_g = \log(\lambda_g), g = 1, \ldots, G$.

For factor loadings $\mathbf{B}$ and variance/covariance matrix $\boldsymbol{\Phi}$ and $\boldsymbol{\Psi}$, we specify the following conjugate prior distributions. For $k = 1, \ldots, p$,

$$\mathbf{b}_k \sim N(\mathbf{b}_{k0}, \psi_k\boldsymbol{\Sigma}_{bk0}), \quad \psi^{-1}_k \sim Gamma(a_{\epsilon 0}, b_{\epsilon 0}), \quad \boldsymbol{\Phi}^{-1} \sim Wishart(\mathbf{R}_0, \rho_0), \tag{7}$$

where ψ_k is the kth diagonal element of $\mathbf{\Psi}$; $\mathbf{b}_k$ is the kth row of $\mathbf{B}$; and $\mathbf{b}_{k0}$, $\mathbf{\Sigma}_{bk0}$, $a_{\epsilon 0}$, $b_{\epsilon 0}$, ρ_0, and $\mathbf{R}_0$ are pre-assigned hyperparameters.

Notably, the prior distributions of $\lambda_g, \mathbf{b}_k, \psi_k^{-1}$, and $\mathbf{\Phi}^{-1}$ are conjugate priors, which lead to manageable posterior distributions (see Web Appendix.3) for efficient posterior inference. The preassigned hyperparameters represent available prior information. For instance, if we have some prior knowledge about $\boldsymbol{\beta}$, then we may assign $\boldsymbol{\mu}_\beta$ in accordance with the prior knowledge and σ_β to a small value to reflect accurate information. If the prior information of $\boldsymbol{\beta}$ is unavailable, then we may assign $\boldsymbol{\mu}_\beta$ to an ad hoc value (eg, $\mathbf{0}$) and σ_β to a large value to reflect vague information. The hyperparameters involved in other prior distributions can be assigned in a similar manner. Detailed discussion can be found in Song and Lee.[23]

The Bayesian estimate of $\boldsymbol{\theta}$ can be obtained through the mean or mode of the posterior samples drawn from $p(\boldsymbol{\theta}|\mathbf{Y})$. However, direct sampling from $p(\boldsymbol{\theta}|\mathbf{Y})$ is intractable because of the existence of latent variables. Instead, we work on $p(\boldsymbol{\theta}, \mathbf{U}|\mathbf{Y})$ and use the Gibbs sampler to iteratively simulate each unknown from its full conditional distribution. The details of the posterior inference are provided in the Web Appendix.

3.2 | Dynamic prediction

Suppose a new subject i comes to the clinic at time t with a set of past multivariate longitudinal measurements $\mathcal{Y}_i^{\{t\}} = \{\mathbf{y}_{ij}; j = 1, \ldots, m_i, t_{im_i} \leq t\}$ and baseline covariates $\mathbf{z}_i$. We focus on a time frame $(t, t+\delta]$, within which an intervention for improving subjects' survival is available. Let $\mathcal{D}_n = \{V_i, \Delta_i, \mathbf{y}_i, \mathbf{z}_i; i = 1, \ldots, n\}$ denote the sample, based on which the proposed model is fitted. The conditional probability of survival time $t+\delta$ given survival up to t can be calculated as follows:

$$\begin{aligned}
\pi_i(t+\delta|t) &= \Pr\left(T_i \geq t+\delta | T_i > t, \mathcal{Y}_i^{\{t\}}, \mathbf{z}_i, \mathcal{D}_n\right) \\
&= \int\int \Pr\left(T_i \geq t+\delta | T_i > t, \mathcal{Y}_i^{\{t\}}, \mathbf{z}_i, \mathbf{u}_i, \mathcal{D}_n; \boldsymbol{\theta}\right) p\left(\mathbf{u}_i | T_i > t, \mathcal{Y}_i^{\{t\}}, \mathbf{z}_i, \mathcal{D}_n; \boldsymbol{\theta}\right) p(\boldsymbol{\theta}|\mathcal{D}_n) d\mathbf{u}_i d\boldsymbol{\theta} \\
&= \int\int \frac{\Pr(T_i \geq t+\delta | \mathbf{z}_i, \mathbf{u}_i, \mathcal{D}_n; \boldsymbol{\theta})}{\Pr(T_i \geq t | \mathbf{z}_i, \mathbf{u}_i, \mathcal{D}_n; \boldsymbol{\theta})} p\left(\mathbf{u}_i | T_i > t, \mathcal{Y}_i^{\{t\}}, \mathbf{z}_i, \mathcal{D}_n; \boldsymbol{\theta}\right) p(\boldsymbol{\theta}|\mathcal{D}_n) d\mathbf{u}_i d\boldsymbol{\theta},
\end{aligned} \tag{8}$$

where $\Pr(T_i \geq t | \mathbf{z}_i, \mathbf{u}_i, \mathcal{D}_n; \boldsymbol{\theta}) = \exp\left\{-\sum_{g=1}^{G} v_{ig} \sum_{k=1}^{g} \int_{s_{k-1}}^{\min(s_k, t)} \lambda_k \exp\left[\boldsymbol{\beta}^T \mathbf{z}_i + \alpha h_i(s)\right] ds\right\}$, and $p(\boldsymbol{\theta}|\mathcal{D}_n)$ is the posterior distribution of the parameters given the observed data $\mathcal{D}_n$. Here, the sample size n is assumed to be sufficiently large, so that the standard asymptotic Bayesian theory[30(sect.10.6)] can be applied, and $p(\boldsymbol{\theta}|\mathcal{D}_n)$ can then be approximated by $N(\hat{\boldsymbol{\theta}}, \hat{\mathcal{H}})$ with $\hat{\boldsymbol{\theta}}$ being the Bayesian estimate of $\boldsymbol{\theta}$ and $\hat{\mathcal{H}} = \hat{\text{var}}(\hat{\boldsymbol{\theta}})$. Consequently, a Monte Carlo estimate of $\pi_i(t+\delta|t)$ can be obtained as follows:

$$\hat{\pi}_i(t+\delta|t) = \frac{1}{J} \sum_{j=1}^{J} \frac{\Pr\left(T_i^* \geq t+\delta | \mathbf{z}_i, \mathbf{u}_i^{(j)}, \mathcal{D}_n; \boldsymbol{\theta}^{(j)}\right)}{\Pr\left(T_i^* \geq t | \mathbf{z}_i, \mathbf{u}_i^{(j)}, \mathcal{D}_n; \boldsymbol{\theta}^{(j)}\right)}, \tag{9}$$

where $\boldsymbol{\theta}^{(j)} \sim N(\hat{\boldsymbol{\theta}}, \hat{\mathcal{H}})$, $\mathbf{u}_i^{(j)} = (u_{i1}^{(j)}, \ldots, u_{iL}^{(j)})^T \sim p(\mathbf{u}_i | T_i > t, \mathcal{Y}_i^{\{t\}}, \mathbf{z}_i, \mathcal{D}_n; \boldsymbol{\theta})$, and J is the number of Monte Carlo samples. Similarly, latent variable $\eta_{i,t+\delta}$ can be estimated by

$$\hat{\eta}_{i,t+\delta} = \frac{1}{J} \sum_{j=1}^{J} \sum_{l=1}^{L} u_{il}^{(j)} W_l(t+\delta). \tag{10}$$

Other prediction results, such as predicted SEs and credible intervals, can be obtained in a similar manner by calculating the summaries of the Monte Carlo samples.

The prediction of survival probability and latent cognitive impairment are not static and can be dynamically updated based on the newest measurements of cognitive tests. Suppose that subject i has not suffered from AD by time $t+\delta$. Then, the histories of cognitive measurements are updated to $\mathcal{Y}_i^{\{t+\delta\}} = \{\mathbf{y}_{ij}; j = 1, \ldots, m_i, t_{im_i} \leq t+\delta\}$. Hence, we can update the posterior distribution to $p(\mathbf{u}_i | T_i > t+\delta, \mathcal{Y}_i^{\{t+\delta\}}, \mathbf{z}_i, \mathcal{D}_n; \boldsymbol{\theta})$ and obtain the updated predictive survival probability

$\hat{\pi}_i(t+\delta'|t+\delta)$ and latent cognitive impairment $\hat{\eta}_{i,t+\delta'}$ with $\delta' > \delta$. The performance of the proposed dynamic prediction procedure is evaluated by the time-dependent integrated area under the receiver operating characteristic curve (AUC) and Pearson correlation. Specifically, AUC measures how well the proposed model can discriminate patients who will experience the onset of AD from patients who will not.[7,31,32] Given a random pair of subjects (i_1, i_2), the AUC value can be calculated as follows:

$$\mathrm{AUC}(t,\delta) = P(\pi_{i_1}(t+\delta|t) < \pi_{i_2}(t+\delta|t)|\{T_{i_1} \in (t, t+\delta]\} \cap \{T_{i_2} > t+\delta\}). \tag{11}$$

If subject i_1 suffers from AD within the time frame $(t, t+\delta]$, whereas subject i_2 does not, a good prediction model is expected to assign a higher survival probability for subject i_2, thereby resulting in a higher AUC value. Given that subjects' time-to-AD are not fully observed in the ADNI dataset, we follow Andrinopoulou et al[7] to assign additional weights for the pairs of subjects who cannot be compared due to censoring. The Pearson correlation, which assesses the prediction accuracy of cognitive impairment, is calculated between the predicted latent cognitive impairment $\hat{\eta}_{it}$ and its true value η_{it}.[33] However, η_{it} is indirectly observable. Thus, we first obtain its estimate $\tilde{\eta}_{it}$ based on a separate CFA model (1) and then calculate the Pearson correlation between $\hat{\eta}_{it}$ and $\tilde{\eta}_{it}$.

Notably, some machine learning techniques, such as gradient boosting and Bayesian additive regression tree, may provide comparable or even higher prediction accuracy in the ADNI study. However, compared with the proposed statistical model, such kinds of black-box prediction procedures may not be preferable in the clinical trial of AD due to its loss of interpretability and low stability. If an opaque machine learning method is adopted, the scientific findings that reveal the effects (directions and magnitudes) of risk factors on AD hazards remain completely hidden because the model only gives prediction without explanations. By contrast, a statistical model enables doctors/neuroscientists to discover the mechanism of AD progression and how important risk factors affect AD hazards, thereby facilitating early diagnosis or efficient prevention of AD. Furthermore, patients tend to trust and accept the prediction results of a statistical model with justifiable analyses rather than those of a black-box prediction. This is essential for patients to better cooperate with doctors during treatment.

4 | ALZHEIMER'S DISEASE NEUROIMAGING INITIATIVE DATA ANALYSIS

4.1 | Alzheimer's disease neuroimaging initiative

ADNI was first launched in 2004 under the leadership of Dr. Michael W. Weiner. It was funded as a private-public partnership with $67 million donated by 20 companies, the Foundation for the National Institute, and the National Institute of Aging. The first 5-year study (ADNI-1) was extended by 2 years in 2009 by a Grand Opportunities grant (ADNI-GO), and in 2011 and 2016 by further competitive renewals of the ADNI-1 grant (ADNI-2 and ADNI-3, respectively). The overarching objective of ADNI is to detect AD at the earliest possible stage and identify ways to track the disease progression with biomarkers. On the basis of this objective, ADNI recruited participants between the ages of 55 and 90 years old at 57 cities in the United States and Canada and collected their imaging, genetic, clinical, and cognitive data at baseline, 3, 6, 12, 18, 24 months and every following 12 months during the cohort study. The first phase of ADNI (ADNI-1) aimed to enroll 800 adults, including approximately 200 people for elderly controls, 400 people with MCI, and 200 people with early AD. Subsequently, the three extensions, ADNI-GO, ADNI-2, and ADNI-3, kept the existing participants who enrolled in the former study and further recruited 200, 450, and 371 new participants, respectively, into the cohort. For up-to-date information about ADNI, see adni.loni.usc.edu.

4.2 | Data description

We focused on 715 patients who enrolled in ADNI-1, ADNI-2, or ADNI-3 and suffered from MCI at baseline. Among the 715 MCI patients, the number of follow-up visits varied from 4 to 14, and the maximum visiting time is 132 months.

We considered the longitudinal scores of ADAS11, ADAS13, and FAQ over time for each MCI patient. ADAS is a neuropsychological assessment that tests written and verbal responses of subjects. These responses are related to fundamental cognitive functions. The total score is usually reported as a composite scale of 11 (ADAS11) or expanded to 13

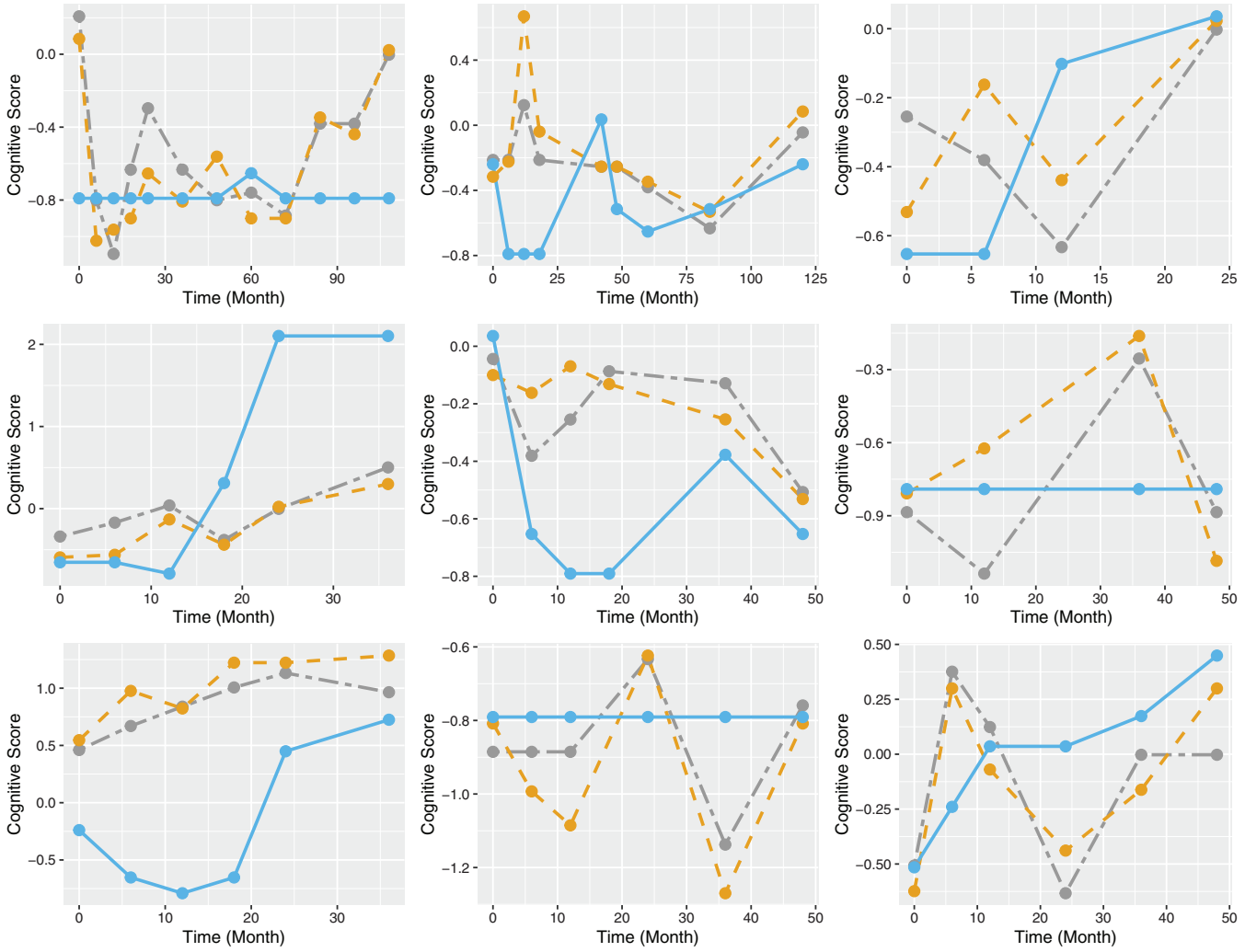

FIGURE 1 Longitudinal measurement of ADAS11, ADAS13, and FAQ of 9 patients in the ADNI study. Grey dot-dash line: ADAS11 scores; yellow dashed line: ADAS13 scores; blue solid line: FAQ scores [Colour figure can be viewed at wileyonlinelibrary.com]

items (ADAS13) and ranges from 0 to 70 or 85. A high score indicates poor cognitive ability. Specific tasks in ADAS11 include word recall, naming objects and fingers, commands, constructional praxis, orientation, word recognition, and language. However, some other cognitive domains, such as attention and concentration, planning and executive function, and praxis, are identified as important treatment targets of anti-dementia drugs but are not assessed by the ADAS11.[34] Such issues are additionally addressed by ADAS13 through a test of delayed word recall and a number cancellation or maze task. By contrast, FAQ is a functional and behavioral assessment with 10 items. It tests a participant's capability to perform daily life tasks in multiple domains, such as visuospatial abilities, planning, organization, and divided attention. FAQ score ranges from 0 to 30, with high scores reflecting great functional dependence. Although ADAS11, ADAS13, and FAQ all assess individual's cognitive impairment, they reflect the cognitive dysfunction from different aspects and do not replace one another. Figure 1 illustrates the observed longitudinal scores of these three cognitive assessments from 9 participants. Apparently, if one uses ADAS11 only to measure cognitive impairment, then the insufficient information cannot reveal the risk that is reflected by abnormally high scores of FAQ and ADAS13, leading to incorrect conclusion for patient with normal neuropsychological function but severe dysfunction in daily activities.

According to the group specific inclusion criteria shown in the general procedures manual of the ADNI study, subjects who met all the following conditions was diagnosed clinically as AD: (i) memory complaint by subject or study partner that is verified by a study partner; (ii) abnormal memory function documented by scoring below the education adjusted cutoff on the Logical Memory II subscale from the Wechsler Memory Scale—revised; (iii) Mini-Mental State Examination (MMSE) between 20 and 26; (iv) Clinical Dementia Rating = 0.5; and (v) NINCDS/ADRDA criteria for probable AD.

Notably, in the ADNI dataset, a total of four cognitive measures, such as ADAS11, ADAS13, FAQ, and MMSE, are available. Considering that MMSE was adopted in the diagnosis of AD (the outcome of interest), we excluded it and used all the remaining measures to characterize cognitive impairment to avoid circularity. Among the 715 MCI patients in the current study, only 294 have progressed to AD during the cohort study and 421 remained at the MCI stage. The time to conversion to AD was the period from the baseline to the date of the first diagnosis of AD or the date of the last visit, whichever came first. The censoring rate of AD was approximately 59%.

Besides the multivariate longitudinal and time-to-event data, we also included patient's clinical and genetic variables at baseline. The clinical characteristics include gender (0 = male , 1 = female), marital status (0 = not married, 1 = married), and education level. The apolipoprotein E (APOE)-ϵ4, a risk factor for early onset AD,[35] is included as a genetic covariate. In the demographic information, 424 individuals are male and 191 are females. Among these individuals, 276 carry one APOE-ϵ4 allele, 80 carry two APOE-ϵ4 alleles, and 359 are non-carriers. For marital status, 558 were married and 157 were not married (divorced, widowed, or never married) at baseline. On average, the patients have 16.0 years of education with a SD of 2.8 years.

4.3 | Data analysis

We considered the proposed model defined by (1)-(3) with $p = 3$ and $q = 1$. Then, we grouped ADAS11 ($y_{ij,1}$), ADAS13 ($y_{ij,2}$), and FAQ ($y_{ij,3}$) into a latent variable "cognitive impairment (η_{ij})" for $i = 1, \ldots, n; j = 1, \ldots, m_i$ as follows:

$$\mathbf{y}_{ij} = \mathbf{B}\eta_{ij} + \boldsymbol{\epsilon}_{ij}, \qquad \mathbf{B}^T = (1^* \; b_2 \; b_3), \tag{12}$$

where the first element of $\mathbf{B}$ was fixed at 1 to obtain an identified model. The trajectory of individual's cognitive impairment is further described as follows:

$$\eta_{ij} = h_i(t_{ij}) = \sum_{l=1}^{L} u_{il} W_l(t_{ij}). \tag{13}$$

In this study, we used a simple version of natural cubic spline derived from a truncated power series basis function[11] to approximate the trajectories: $W_1(t_{ij}) = 1, W_2(t_{ij}) = t_{ij}$, and $W_{l+2}(t_{ij}) = e_l(t_{ij}) - e_{L-1}(t_{ij})$ for $l = 1, \ldots, L-2$, where $e_l(t_{ij}) = [(t_{ij} - \kappa_l)_+^3 - (t_{ij} - \kappa_L)_+^3]/(\kappa_L - \kappa_l)$, and $\kappa_l, l = 1, \ldots, L$ are the knots taken in the range of t_{ij}. We chose 6 most-visited follow-up time (more than 200 subjects recorded in this visit), 6, 12, 18, 24, 36, and 48 months, as the knots ($L = 6$) in this study. For the PH model, we included time-invariant covariates, namely, marital status (z_{i1}), gender (z_{i2}), education level (z_{i3}), and APOE-ϵ4 genetic covariate coded using two dummy variables: "one APOE-ϵ4 allele carrier (z_{i4})" and "two APOE-ϵ4 alleles carrier (z_{i5})" at baseline as well as time-variant covariate cognitive impairment $h_i(t)$ as follows:

$$\lambda(t|\mathbf{z}_i, \mathbf{u}_i) = \lambda_0(t) \exp\left(\sum_{r=1}^{5} \beta_r z_{ir} + \alpha h_i(t)\right). \tag{14}$$

Let $t_{ij} = \tilde{t}_{ij}/100$ for the purpose of reducing the scale of t_{ij}. The continuous variables $y_{ij,1}, y_{ij,2}, y_{ij,3}$, and z_{i3} were standardized prior to analysis. We used $G = 10$ sub-intervals to model the piecewise baseline hazard function. The cut points $s_0, \ldots, s_G$ that define these sub-intervals were set as the quantiles of the observed survival times. The prior distributions in (6) and (7) with prior inputs were adopted as follows: $\boldsymbol{\mu}_\beta = \mathbf{0}$, $\mu_\alpha = 0, \sigma_\beta^2 = \sigma_\alpha^2 = 1, \mathbf{b}_{10} = \mathbf{0}, \boldsymbol{\Sigma}_{b10} = \mathbf{I}_3, a_{\epsilon 0} = 9, b_{\epsilon 0} = 4, \rho_0 = 7, \mathbf{R}_0 = 4\mathbf{I}_6, a_1 = 2$, and $a_2 = 0.01$.

We ran several parallel sequences from different starting values of parameters to check convergence of the MCMC algorithm and found that the algorithm converged within 30 000 iterations. Thus, we collected 30 000 posterior observations after 30 000 burn-in iterations to conduct Bayesian inference. Table 1 presents the Bayesian estimates together with their SE estimates (in parentheses) for all the parameters. The Bayesian estimates of baseline hazard $\lambda_g, g = 1, \ldots, 10$ are given in Table S5 of Web Appendix. Several findings were obtained as follows:

First, the estimated factor loadings are $\hat{b}_2 = 1.014$ (0.007) and $\hat{b}_3 = 0.773$ (0.012), which imply strong association between the latent variable (cognitive impairment) and it's observed indicators (ADAS11, ADAS13, and FAQ). Furthermore, we used two indices, namely, construct reliability and average variance extracted (AVE),[36] to check the

TABLE 1 Parameter estimates in the analysis of ADNI data

M_p: Joint model with latent variables			M_{ind}: Conventional joint model		
Variable	**Est**	**SD**	**Variable**	**Est**	**SD**
Marital status(β_1)	−0.337*	0.149	Marital status(β_1)	−0.291	0.185
Gender(β_2)	0.234	0.130	Gender(β_2)	−0.046	0.156
Education(β_3)	−0.041	0.057	Education(β_3)	0.006	0.067
One APOE-ϵ4 allele (β_4)	0.073	0.126	One APOE-ϵ4 allele (β_4)	−0.073	0.155
Two APOE-ϵ4 alleles (β_5)	0.364*	0.170	Two APOE-ϵ4 alleles (β_5)	0.161	0.213
			ADAS11 (α_1)	−2.084*	0.341
Cognitive impairment (α)	1.420*	0.068	ADAS13 (α_2)	3.197*	0.390
			FAQ (α_3)	1.209*	0.108
b_2	1.014*	0.007			
b_3	0.773*	0.012			
ψ_1	0.109*	0.003			
ψ_2	0.083*	0.002			
ψ_3	0.466*	0.010			
φ_{11}	0.348*	0.020			
φ_{22}	5.472*	0.356			
φ_{33}	24.129*	2.420			
φ_{44}	0.508*	0.177			
φ_{55}	0.470*	0.167			
φ_{66}	0.528*	0.182			

Abbreviation: SD, standard deviation.

*Indicates significant parameters under 0.05 significance level.

reliability and validity of these three measures. The construct reliability is defined as $\frac{(\sum_{k=1}^{3} b_k)^2}{(\sum_{k=1}^{3} b_k)^2+\sum_{k=1}^{3}\psi_k}$, and AVE is defined as $\frac{\sum_{k=1}^{3} b_k^2}{\sum_{k=1}^{3} b_k^2+\sum_{k=1}^{3}\psi_k}$, where b_k is the factor loading of the kth measure and ψ_k is the variance of the associated error term. The construct reliability and AVE of the proposed model are 0.97 and 0.92, respectively, which far exceed the commonly used thresholds 0.7 and 0.5. Therefore, the three measures are valid and reliable in measuring the cognitive impairment.

Second, marital status [$\hat{\beta}_1 = -0.337(0.149)$] exerts a significant negative effect on AD hazards, thereby indicating that married people have a lower risk of developing AD than unmarried people. This finding is in line with the published reports.[37] On the contrary, the effect of two APOE-ϵ4 alleles [$\hat{\beta}_5 = 0.364(0.170)$] on cognitive impairment is significantly positive, implying that two APOE-ϵ4 alleles carriers are more likely to develop AD than one or none APOE-ϵ4 allele carriers. This result agrees with several published reports[38,39] and implies that APOE-ϵ4 alleles are important risk factors in cognitive dysfunction. In addition, gender [$\hat{\beta}_2 = 0.234(0.130)$] has a marginally significant positive effect on AD hazards because its 95% credible interval includes zero but its 90% credible interval ([0.031, 0.466]) does not include zero. Hence, females have a relatively higher risk of developing AD than males. This finding is in line with those of published reports[40,41] and may be related to the fact that the advantages of mitochondria, such as protecting against amyloid-β toxicity, generating less reactive oxygen species, and releasing less apoptogenic signals, are eventually lost for old women. The other covariates, such as education level and carrying one APOE-ϵ4 allele, hardly influence AD hazards when marital status, gender, two APOE-ϵ4 alleles, and the individual trajectories of cognitive impairment are controlled.

Third, the estimates of elements in the covariance matrix $\mathbf{\Phi}$ of the random effect are significant, which confirms the existence of heterogeneity among patients' cognitive decline. Figure 2 presents subjects' estimated trajectories of longitudinal cognitive impairment during MCI phase based on the posterior means of the random effects. Apparently, the cognitive impairment exhibits more pronounced ascending trend for patients who would later convert to AD ($\Delta_i = 1$)

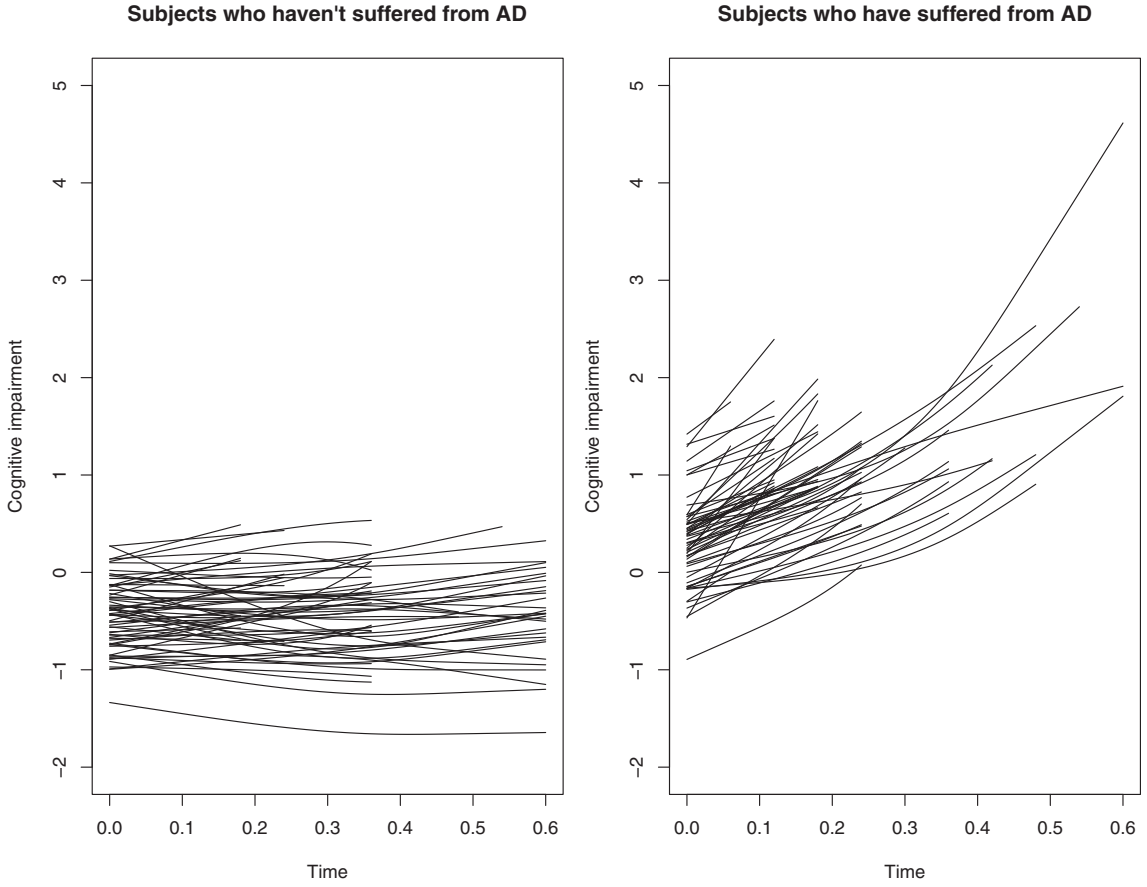

FIGURE 2 Patients' estimated trajectories of longitudinal cognitive impairment during MCI phase. Left panel: patients who would stay at the MCI stage ($\Delta_i = 0$); right panel: patients who would later convert to AD ($\Delta_i = 1$)

than for patients who would not ($\Delta_i = 0$). Although previous studies (eg, Chen et al,[42] Cloutier et al[43]) have reported that characterizing the pattern of cognitive decline during MCI phase is crucial to early diagnosis of AD, their results are based only on a single cognitive measure and thus fail to reveal the overall pattern and effect of multiple measures of cognitive impairment in MCI-AD conversion. By contrast, our obtained result provides a potential of introducing the trajectory function of an integrated measure of cognitive impairment as a marker of AD progression and is therefore beneficial to early diagnosis of AD. When new drugs for AD treatment are available, such integrated cognitive measure may also be used for selection of cases to evaluate drug efficiency in slowing the disease progression.

Fourth, for most of the patients (over 81%) who had suffered from AD, the posterior means of u_{i2} and u_{i3} are significantly positive. Such random effects correspond to the second and third basis splines, namely, $W_2(t) = t$ and $W_3(t) = [(t-\kappa_1)_+^3 - (t-\kappa_6)_+^3]/(\kappa_6-\kappa_1) - [(t-\kappa_5)_+^3 - (t-\kappa_6)_+^3]/(\kappa_6-\kappa_5)$, where $W_3(t) > 0$ if $t > \kappa_1$ and $W_3(t) = 0$ if $0 \le t \le \kappa_1$. This result indicates that patients' cognitive impairment shows linear deteriorating trends from baseline to 6 months ($\kappa_1 = 6/100 = 0.06$), but nonlinear and more pronounced deteriorating trends after 6 months. This finding may elicit a highly targeted treatment in AD prevention. The current treatment for mild to moderate cognitive impairment includes medications for memory loss, such as cholinesterase inhibitors,[44] and non-drug approaches for behaviors, such as physical, emotional, and social stimulation. The obtained result provides evidences of an accelerate disease deterioration over time and suggests an increase in drug doses or non-drug treatment, such as the validation therapy[45] provided by caregivers for patients who have suffered from MCI for more than 6 months.

Fifth, the trajectory function of cognitive impairment exhibits a significant positive effect on AD hazards [$\hat{\alpha} = 1.420(0.068)$], which implies that people with severe cognitive impairment are at high risk of developing AD. Furthermore, the strong association between cognitive impairment and AD hazards confirms that a comprehensive measure of cognitive impairment can be used as a pre-diagnosis for AD.

To assess how the above results are sensitive to the spline parameters, we let the number of knots in the spline approximation vary from 3 to 9 and changed the spline basis function from natural cubic splines to cubic Hermite splines, of which each piece is a third-degree polynomial specified in Hermite form. The knots were chosen as the first l elements in $\{6, 12, 18, 24, 36, 48, 60, 72, 84\}$ months, for $l = 3, \ldots, 9$. We reanalyzed the ADNI dataset using the disturbed number of knots and basis functions. The nonlinear patterns of cognitive impairment trajectories and other results obtained using different spline parameters are similar to those presented above and not reported. Moreover, we conducted sensitivity analysis with ad hoc disturbances to the current prior inputs and found that the Bayesian estimates (see Figure S4 of Web Appendix) were robust to the prior inputs under consideration.

4.4 | Model comparison

To check the necessity of a nonlinear trajectory model, we reanalyzed the ADNI dataset by using a simpler joint model with linear trajectory, denoted as M_{linear}. Figure 3 presents the estimated and true dynamic cognitive impairment trajectories of 9 randomly selected patients. The estimated trajectories of η_{ij} based on the proposed model, denoted as M_{p}, are reasonably close to those obtained from a separate factor analysis, indicating the satisfactory performance of the proposed joint model in describing the complex individual trajectories of cognitive decline. By contrast, M_{linear} that assumes a linear trajectory of η_{ij} cannot reveal the nonlinear patterns of patients' cognitive impairment. Furthermore, we compared the out-of-sample prediction performances of the two models. We focused on the time frame with $t = 12, 18$ months and $\delta = 6, 24$ months in predicting the survival probability $\hat{\pi}(t + \delta|t)$ and chose $t = 12$ months and $\delta = 6, 12, 24$ months in predicting the latent cognitive impairment $\hat{\eta}_{i,t+\delta}$. The full dataset was randomly split into a training set with 515 subjects and a test set with 200 subjects. The random split was repeated 100 times. For each split, we fitted M_p and M_{linear} to the training set and then calculated AUC and Pearson correlation based on the test set. Table 2 (upper panel) presents the AUC values for M_{p} and M_{linear}. The AUC values of M_{p} are consistently higher than those of M_{linear} in all combinations of t and δ, indicating that the proposed joint model with flexible nonlinear trajectory has better discriminative capability than the simpler model M_{linear} in the diagnosis of AD. In addition, the Pearson correlations of $\hat{\eta}_{ij}$ and $\tilde{\eta}_{ij}$ calculated based on M_p at $\delta = 6, 12$, and 24 months are 0.862, 0.802, and 0.713, respectively, whereas those of M_{linear} are 0.856, 0.788, and 0.691, respectively. The proposed model achieves higher prediction accuracy in all scenarios, which reconfirms the superiority of the flexible nonlinear trajectory in describing the complex individualized dynamic pattern of cognitive impairment.

To investigate the necessity of latent variable, we reanalyzed the dataset using a conventional joint model by considering three independent trajectory models for ADAS11 ($y_{ij,1}$), ADAS13 ($y_{ij,2}$), and FAQ ($y_{ij,3}$). In this conventional model, denoted by M_{ind}, instead of grouping the three longitudinal measurements into a latent variable η_{ij} and then examining the overall effect of the trajectory of η_{ij} on the hazard function, we independently modeled the individual trajectories of $y_{ij,1}, y_{ij,2}$, and $y_{ij,3}$ and then examined their separate effects on the hazard function, as follows:

$$\lambda(t|\mathbf{z}_i, \mathbf{u}_i) = \lambda_0(t) \exp\left(\sum_{r=1}^{5} \beta_r z_{ir} + \alpha_1 h_{1i}(t) + \alpha_2 h_{2i}(t) + \alpha_3 h_{3i}(t)\right), \tag{15}$$

where $h_{1i}(t), h_{2i}(t)$, and $h_{3i}(t)$ are the trajectories of $y_{ij,1}, y_{ij,2}$, and $y_{ij,3}$, respectively, as described by Model (2). The DIC values of M_p, M_{linear}, and M_{ind} are 37914.84, 38292.97, and 35520.77, respectively. Although M_{ind} shows the lowest DIC value, it yields misleading predictor effects due to multicollinearity. Table 1 provides a comparison of the estimation results obtained using the proposed joint model and M_{ind} without latent variable [see (15)]. The estimated effects of marital status, two APOE-ϵ4 alleles, and gender were significant or marginally significant in M_{p} but became insignificant in M_{ind}. Moreover, the effect of ADAS11 on AD hazards was reversed [$\hat{\alpha}_1 = -2.084(0.341)$], implying that high ADAS11 scores (severe cognitive impairment) were associated with low AD hazards. Thus, M_{ind} overlooked the important risk factors of AD and produced contradictory results. Further verification showed that the pairwise sample correlations among the scores of ADAS11, ADAS13, and FAQ were 0.977, 0.669, and 0.689. Such high correlations induced the multicollinearity problem and misleading results of M_{ind}. We also conducted the out-of-sample prediction to compare M_{p} and M_{ind}. Table 2 (upper panel) presents their AUC values under $t = 12$, 18 months and $\delta = 6$, 24 months. The AUC values of M_{p} are generally greater than those of M_{ind}, indicating that the proposed model possesses higher prediction capacity than M_{ind}. The only exception occurs in the time frame [$t = 18$ month, $\delta = 6$ month], where the AUC value of M_{p} is slightly lower than that of M_{ind}. This phenomenon was also reported by existing literature (eg, Kutner et al,[46] Weiss[47]), which revealed

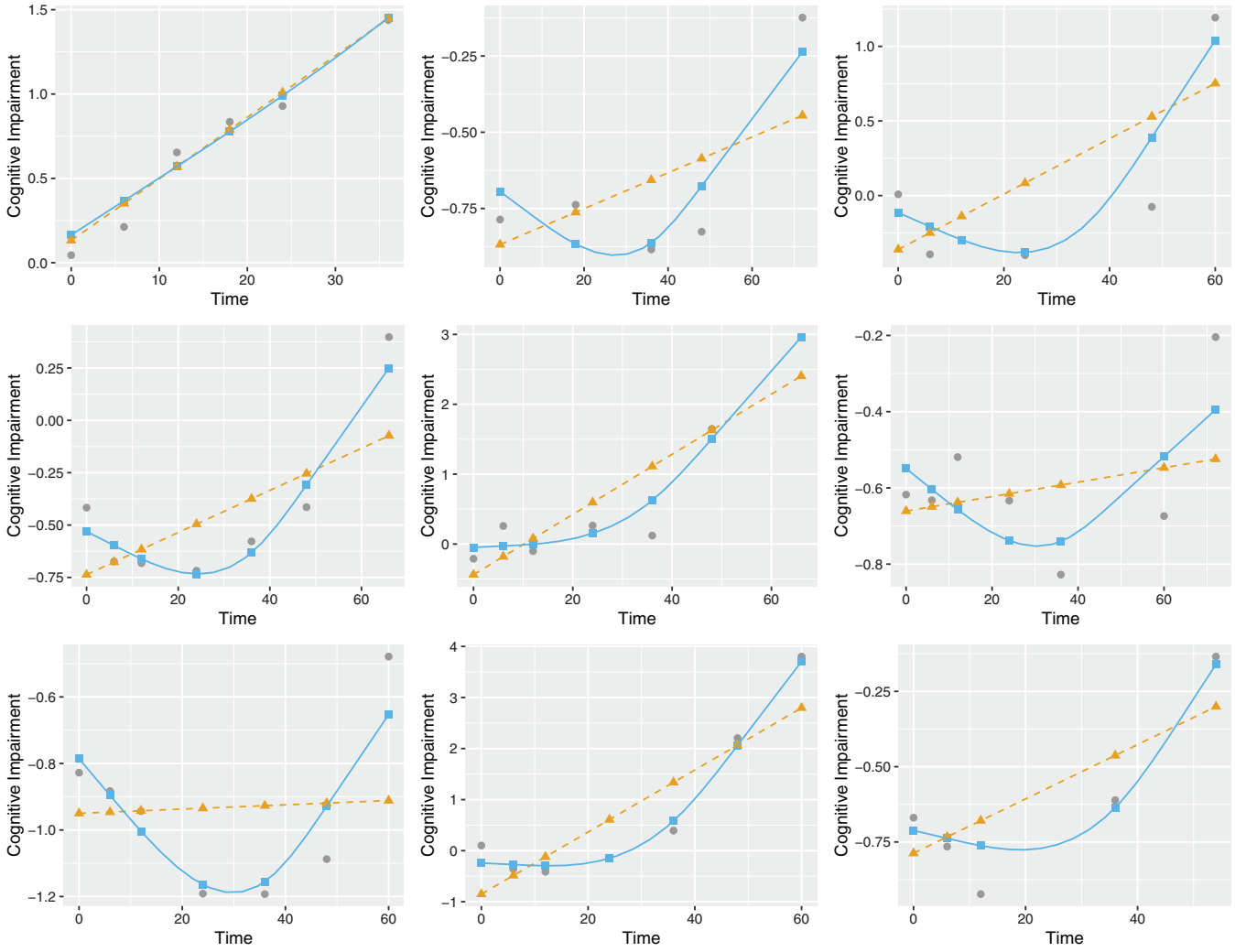

FIGURE 3 Expected cognitive impairment of 9 randomly selected patients at each visit. Grey dot: the values of η_{ij} calculated using a separate CFA model; blue solid curve: cognitive impairment trajectory estimated using the proposed model M_p; yellow dashed curve: cognitive impairment trajectory estimated using a simpler model M_{linear} with linear trajectory [Colour figure can be viewed at wileyonlinelibrary.com]

TABLE 2 AUC values of competing models in the ADNI study

	$t = 12$ month		$t = 18$ month	
Model	$\delta = 6$ month	$\delta = 24$ month	$\delta = 6$ month	$\delta = 24$ month
M_{p}	0.851	0.841	0.855	0.839
M_{linear}	0.843	0.831	0.852	0.831
M_{ind}	0.847	0.803	0.861	0.803
M_{FULL}	0.835	-	0.840	-
M_{ADAS11}	0.818	-	0.819	-
M_{ADAS13}	0.826	-	0.828	-
M_{FAQ}	0.781	-	0.825	-

that the presence of multicollinearity in predictors may not affect the overall fit of the model and its ability to accurately predict the response value in certain cases, but it may sap the significant of some predictors and change their signs, thereby making the specification of the correct model difficult. Nevertheless, the conventional model M_{ind} is not preferable even in this scenario due to its loss of interpretability, which is essential for doctors to understand the important risk factors of the AD progression.

We also examined the preponderance of the combined measure of cognitive impairment over a single cognitive measure. We focused on 57 patients with diverse patterns of different cognitive measures. For instance, some patients' ADAS11 scores exhibited an increasing trend but their FAQ scores remained unchanged (see Figure 1). For these patients, any single cognitive test failed to capture the full picture of their cognitive impairment. These 57 subjects were treated as a test set and the remaining 658 subjects were treated as a training set. We compared the proposed model M_{FULL} with three reduced models M_{ADAS11}, M_{ADAS13}, and M_{FAQ}, each of which used a single measure to assess cognitive impairment. Given that most of the subjects in the test set had suffered from AD or were censored at the 24th month, we calculated AUC at $t = 12, 18$ months and $\delta = 6$ months. Table 2 (lower panel) presents the AUC values of the four competing models. M_{FULL} consistently outperforms the three reduced models. Hence, the use of a combined measure of cognitive impairment enhances the discriminative capability of the model.

5 | SIMULATION STUDY

In this section, we conducted two simulation studies to evaluate the finite sample performance of the proposed model. Simulation 1 shows the empirical performance of the Bayesian estimation, whereas Simulation 2 examines the sensitivity of the inference results to model assumptions.

5.1 | Simulation 1

We generated datasets based on the proposed model with $i = 1, \dots, n$, time points m_i varying from 6 to 9, $p = 3$, and $q = 1$. Model (1) was defined as follows:

$$\mathbf{y}_{ij} = \mathbf{B}\eta_{ij} + \boldsymbol{\epsilon}_{ij}, \quad \mathbf{B}^T = (1^* \; b_2 \; b_3), \tag{16}$$

where the first elements of $\mathbf{B}$ was fixed as 1 to obtain an identified model, $b_2 = 0.6$, and $b_3 = 0.7$. The random error $\boldsymbol{\epsilon}_{ij}$ was generated from $N(\mathbf{0}, \boldsymbol{\Psi})$ with $\boldsymbol{\Psi} = \text{diag}(\psi_1, \psi_2, \psi_3) = \text{diag}(0.3, 0.3, 0.3)$.

The trajectory model (2) was set as follows:

$$\eta_{ij} = h_i(t_{ij}) = \sum_{l=1}^{6} d_{il} W_l(t_{ij}), \tag{17}$$

where $\mathbf{d}_i = (d_{i1}, \dots, d_{i6})^T$ are the subject-specific random effects independently simulated from $N(\mathbf{0}, \mathbf{I}_6)$, and $W_l(t_{ij})$s are natural cubic basis splines.

The PH model (3) was defined as follows:

$$\lambda(t|\mathbf{z}_i, \mathbf{u}_i) = \lambda_0(t) \exp\left(\boldsymbol{\beta}^T \mathbf{z}_i + \alpha h_i(t)\right), \tag{18}$$

where $\mathbf{z}_i = (z_{i1}, z_{i2}, z_{i3})^T$, and z_{i1}, z_{i2}, and z_{i3} were drawn from $Exp(1) - 1$, $N(0, 1)$, and $t(5)$, respectively, in which $Exp(1)$ denotes the exponential distribution with the rate parameter 1, and $t(5)$ denotes the t distribution with a degree of freedom 5. The true population values of the regression parameters were set as $\boldsymbol{\beta}^T = (1, -1, 1)$ and $\alpha = 1$. The failure time T_i was generated from Model (18) with three types of baseline hazard functions: (i) $\lambda_0(t) = 1$ (constant), (ii) $\lambda_0(t) = t + 0.5$ (linear), and (iii) $\lambda_0(t) = t^2 + 0.3$ (nonlinear). The censoring time C_i was independently generated based on a uniform distribution of $U[c_1, c_2]$, where c_1 and c_2 were selected to achieve censoring rates of 30% and 50%, respectively. Two sample sizes $n = 200$ and $n = 500$ were considered.

In the posterior analysis, we used $G = 5$ sub-intervals to model the piecewise baseline hazard function. The cut points $s_0, \dots, s_G$ that define these sub-intervals were set as the quantiles of the observed survival times. The natural cubic spline

TABLE 3 Parameter estimates in the PH model in Simulation 1

			$\lambda_0(t)=1$		$\lambda_0(t)=t+0.3$		$\lambda_0(t)=t^2+0.5$	
n	**CR**	**Para**	**BIAS**	**RMS**	**BIAS**	**RMS**	**BIAS**	**RMS**
200	30%	α	−0.044	0.098	−0.039	0.093	−0.018	0.090
		β_1	−0.060	0.134	−0.074	0.130	−0.057	0.125
		β_2	0.034	0.113	0.053	0.114	0.037	0.109
		β_3	−0.052	0.116	−0.065	0.113	−0.051	0.108
	50%	α	−0.061	0.127	−0.065	0.118	−0.050	0.113
		β_1	−0.075	0.143	−0.090	0.142	−0.071	0.136
		β_2	0.064	0.134	0.075	0.141	0.059	0.135
		β_3	−0.080	0.136	−0.089	0.135	−0.072	0.129
500	30%	α	−0.028	0.068	−0.028	0.062	−0.015	0.058
		β_1	−0.052	0.098	−0.062	0.091	−0.047	0.087
		β_2	0.037	0.081	0.055	0.088	0.042	0.081
		β_3	−0.052	0.098	−0.059	0.089	−0.048	0.087
	50%	α	−0.055	0.093	−0.054	0.086	−0.040	0.080
		β_1	−0.064	0.108	−0.067	0.099	−0.050	0.095
		β_2	0.055	0.100	0.063	0.104	0.050	0.096
		β_3	−0.066	0.111	−0.067	0.098	−0.054	0.094

with 6 knots was again used to model the individual's trajectory $h_i(t)$, and the knots $(\kappa_l, l = 1, \ldots, 6)$ were selected as the lth quantiles of all measurement time t_{ij}. We considered the prior inputs of (6)-(7) as follows:

$$\begin{aligned} \text{Prior (I)} \quad & \boldsymbol{\mu}_\beta = \mathbf{0},\ \mu_\alpha = 0, \sigma_\beta^2 = \sigma_\alpha^2 = 1, \mathbf{b}_{10} = \mathbf{0}, \boldsymbol{\Sigma}_{b10} = \mathbf{I}_3, \\ & a_{\epsilon 0} = 9, b_{\epsilon 0} = 4, \rho_0 = 7, \mathbf{R}_0 = 4\mathbf{I}_6, a_1 = 2, a_2 = 0.01. \end{aligned} \tag{19}$$

Three parallel sequences with different starting values of unknowns were generated to determine the number of burn-in iterations. We collected 20 000 observations after discarding 20 000 burn-in iterations to obtain the Bayesian estimates of the model parameters. We used bias (BIAS) and root mean square error (RMS) between Bayesian estimates and their true population values to assess the empirical performance of the parameter estimates. Table 3 summarizes the estimation results of key parameters on the basis of 100 replicated datasets. The estimate of $\boldsymbol{\Phi}$ is unimportant and not reported. The BIAS and RMS for most of the parameters are close to zero, thereby indicating the satisfactory performance of Bayesian estimation in all the settings under consideration. As expected, performance is improved when sample size increases from $n = 200$ to $n = 500$ or censoring rate (CR) decreases from 50% to 30%. To evaluate the accuracy of the Monte Carlo estimation of survival probability, we compare the value of $\hat{\pi}_i(t+\delta|t)$ with its true value $\pi_i(t+\delta|t)$ with $t = x, x$ and $\delta = x$. The estimates of $\pi_i(t+\delta|t)$ (presented in Table S5 of Web Appendix) are close to their true values for all the combinations of t and δ, indicating a satisfactory accuracy of the proposed Monte Carlo estimation.

For comparison, we considered a competing joint model with the same setting but a linear trajectory model as follows:

$$\eta_{ij} = h_i(t_{ij}) = u_{i1} + u_{i2}t_{ij}. \tag{20}$$

The Bayesian results obtained with $\lambda_0(t) = 1$ and CR = 30% under the proposed spline-based trajectory model and the simpler linear trajectory model are summarized in Table S1 of the Web Appendix. All the Bayesian estimates in the proposed model are reasonably close to their true values, while the estimates of α and ψ show considerable bias in the simpler model. Such inaccuracy also appears in the estimates of latent variable η_{ij} and its trajectory $h_i(t)$ as illustrated in Figure S1 of the Web Appendix. The η_{ij} values of nine randomly selected subjects and their corresponding trajectories

$h_i(t)$s estimated using the spline-based random coefficient model are close to each other, indicating that the proposed model can provide adequate flexibility in describing the complex and subject-specific forms of individual trajectories. By contrast, the simpler model with a linear trajectory cannot capture the complex patterns of individual trajectories and produces biased results.

5.2 | Simulation 2

This section assesses the sensitivity of Bayesian estimates to the inputs of prior distributions and to the violation of the normality assumption of $\mathbf{u}_i$ and $\boldsymbol{\epsilon}_{ij}$.

We first disturbed the hyperparameters as follows:

$$\text{Prior (II)} \quad \boldsymbol{\mu}_\beta = (2,2,2)^T,\ \mu_\alpha = 2, \sigma_\beta^2 = \sigma_\alpha^2 = 10^4, \mathbf{b}_{10} = (2,2,2)^T, \boldsymbol{\Sigma}_{b10} = 10^4\mathbf{I}_3,$$
$$a_{\epsilon 0} = 3, b_{\epsilon 0} = 2, \rho_0 = 4, \mathbf{R}_0 = 2\mathbf{I}_6, a_1 = 4, a_2 = 0.001.$$

The Bayesian results obtained with $\lambda_0(t) = 1$ and CR = 30% under Prior (II) are presented in Table S2 of the Appendix along with those obtained under Prior (I). All the estimates under Priors (I) and (II) are similar. Thus, Bayesian estimation results are insensitive to the given prior inputs.

Furthermore, we investigated the sensitivity of Bayesian results to the normality assumption of $\mathbf{u}_i$ and $\boldsymbol{\epsilon}_{ij}$. The model setup is the same as that of Simulation 1, except that the distributions of $\mathbf{u}_i$ or $\boldsymbol{\epsilon}_{ij}$ are no longer normal. We considered several nonnormal cases as follows: (1) $d_{ij} \sim Gamma(4,2) - 2$; (2) $d_{ij} \sim \frac{1}{3}N(1,0.5) + \frac{2}{3}N(-0.5,0.5)$; (3) $\epsilon_{ijk} \sim Beta(3,1) - \frac{3}{4}$; and (4) $\epsilon_{ijk} \sim \sqrt{0.2}t_{(5)}$. Table S3 of the Appendix presents the estimation results under Cases (1)-(4) with $\lambda_0(t) = 1$, CR = 30%, and $n = 500$. A good agreement is achieved between the results reported in Tables S2 and S3 of the Web Appendix. Thus, the Bayesian estimates of the model parameters are robust to the violation of normality assumption of $\mathbf{u}_i$ and $\boldsymbol{\epsilon}_{ij}$ under consideration.

Finally, we use $\tilde{\boldsymbol{\eta}}_{ij}$ as an approximation of $\boldsymbol{\eta}_{ij}$ in calculating the Pearson correlation $Cor(\boldsymbol{\eta}_{ij}, \hat{\boldsymbol{\eta}}_{ij})$ given that $\boldsymbol{\eta}_{ij}$ is unobservable. To check the accuracy of such approximation, we compare the values of $Cor(\boldsymbol{\eta}_{ij}, \hat{\boldsymbol{\eta}}_{ij})$ and $Cor(\tilde{\boldsymbol{\eta}}_{ij}, \hat{\boldsymbol{\eta}}_{ij})$. Their estimates with $n = 300$, $\lambda_0(t) = 1$ and CR = 30% are 0.741 and 0.743, which indicate that $\tilde{\boldsymbol{\eta}}_{ij}$ is a good approximation of $\boldsymbol{\eta}_{ij}$ in the calculation of $Cor(\boldsymbol{\eta}_{ij}, \hat{\boldsymbol{\eta}}_{ij})$.

The computer code for conducting the preceding analyses is written in R and is freely available at https://github.com/kenerous/JMMLS.

6 | DISCUSSION

A joint model with multivariate longitudinal and survival data was developed to investigate the risk factors of the conversion from MCI to AD and predict the time to onset of AD. We used the factor analytic technique to comprehensively characterize patients' cognitive impairment through multiple assessments of cognitive ability and then revealed its trajectory using a highly flexible spline-based method. Such modeling framework solves two main problems of previous AD studies, that is, the incomplete measure of cognitive impairment and linearity assumption on the trajectory of cognitive impairment. While the results of our study confirmed several previous discoveries, such as the effects of marital status, APOE-ϵ4 alleles, and gender on AD hazards, novel findings were also obtained. On the population level, the cognitive impairment trajectory during MCI phase exhibited more pronounced ascending trend for patients who would later convert to AD than for patients who would not. On the individual level, the complex patterns of individual trajectories of cognitive decline were revealed through the proposed spline-based trajectory model. Moreover, the cognitive impairment of patients with AD progression showed linear deteriorating trends from baseline to 6 months but nonlinear and more pronounced deteriorating trends after 6 months. These findings may facilitate an early diagnosis of AD and highly targeted treatment strategies in AD prevention. We also conducted the out-of-sample prediction to compare the proposed model with several existing models. The results show that the proposed model outperforms the existing models in terms of estimation and prediction.

This study can be extended in several directions. First, the proposed model assumes that the measurement errors in Models (1) and the random coefficients in Model (2) follow the multivariate normal distribution. However, this

normality assumption may not be supported in practice. Although our simulation study shows that the estimates of model parameters are relatively insensitive to several non-normal distributions of $\mathbf{u}_i$ and $\boldsymbol{\epsilon}_{ij}$, relaxing the normality assumption and allowing these distributions to be unspecified are considerable scientific interests. For example, introducing the Dirichlet process prior[48] to model the distributions of the random effects and error terms can significantly enhance model flexibility. Second, this study considered scalar quantities, that is, scores of several cognitive tests, as longitudinal measurements. It would be interesting to consider longitudinal functional data, such as patients' 3D MRI scans assessed over time in the ADNI study and establish the joint modeling of longitudinal functional and survival data to examine the association between brain imaging information and time to onset of AD. Third, we exclude subjects with less than four measurements in this study because too few time points are insufficient to uncover the trajectory of cognitive impairment. Such a selection may lead to a selection bias. However, given that our primary objective is to model the trajectory of time-varying cognitive impairment and its effect on AD hazards, using samples with little dynamic information cannot achieve our goal. A highly sophisticated method for addressing this issue is worthy of further investigation. Fourth, as pointed out by the Associate Editor, the Gold Standard to test AD is autopsy. However, ADNI was first launched in 2004, but autopsy had not been included in consent form until 2016. Even after 2016, very few autopsies were available in the ADNI study due to insufficient effort on obtaining consent and insufficient tracking of participants after study withdrawal. The limited samples restricted the use of autopsy as the test of AD in this study. Nevertheless, developing effective strategies to collect autopsy information can facilitate the use of this Gold Standard in future studies. Finally, this study considered a PH model to link the potential risk factors to time-to-AD. The proportional hazards assumptions can be evaluated through the weighted Schoenfeld residual test.[49] In the ADNI study, we checked the proportional hazards assumption for each predictor and found that one of the predictors, $h_i(t)$, did not pass the test. We also considered an accelerated failure time (AFT) model, which does not require the proportionality assumption and directly examines the predictor effects on the logarithm of time-to-AD. However, the AFT model underperforms the PH model in terms of prediction accuracy. For example, the AUC values of the AFT model with $t = 12, 18$ months and $\delta = 6$ months are 0.818 and 0.821, which are much lower than those of the proposed PH model (0.851 and 0.855). Apart from the AFT model, the additive hazards (AH) model is also a common alternative to the PH model due to its easy interpretation of predictor effects. However, the AH model is much more restrictive than its PH counterpart because it assumes that all predictor effects are linear and their summation (together with the baseline hazards) must be nonnegative. Ensuring such nonnegativity elicits additional issues, especially when the predictors include the trajectory function of a latent factor. Furthermore, all published reports in the ADNI study (eg, Lee et al,[16] Li and Luo,[10] Kong et al[17]) used PH models to examine the potential risk factors of AD hazards. In accordance with the existing literature and based on the aforementioned reasons, we used the PH model for the survival outcome in this study although it may not be an optimal choice. Extending the proposed joint modeling framework to incorporate more flexible survival models, such as transformation models, may further enhance the estimation and prediction accuracy, but it poses new theoretical and computational challenges and requires substantial efforts in the future.

ACKNOWLEDGEMENT

This research was supported GRF Grants (14301819, 14302222) from the Research Grant Council of the Hong Kong Special Administration Region. The authors are thankful to the editor, the associate editor, and two anonymous reviewers for their valuable comments and suggestions which improved the article substantially.

DATA AVAILABILITY STATEMENT

Data sharing not applicable—no new data generated

ORCID

Deng Pan https://orcid.org/0000-0003-2785-4666
Xinyuan Song https://orcid.org/0000-0002-4877-3200

REFERENCES

1. Tsiatis AA, Dafni U, DeGruttola V, Propert KJ, Strawderman RL, Wulfsohn M. The relationship of CD4 counts over time to survival in patients with AIDS: Is CD4 a good surrogate marker? In: Jewell NP, Dietz K, Farewell VT, eds. *AIDS Epidemiology*. Vol 73. Boston, MA: Birkhäuser; 1992:256-274.
2. de Gruttola V, Tu X. Modelling progression of CD4-lymphocyte count and its relationship to survival time. *Biometrics*. 1994;50:1003-1014.
3. Cox D. Regression models and life-tables. *J Royal Stat Soc Ser B*. 1972;34:187-220.

4. Brown ER, Ibrahim JG. A Bayesian semiparametric joint hierarchical model for longitudinal and survival data. *Biometrics*. 2003;59:221-228.
5. Ibrahim JG, Chen MH, Sinha DY. Bayesian methods for joint modeling of longitudinal and survival data with applications to cancer vaccine trials. *Stat Sin*. 2004;14:863-883.
6. Barrett J, Su L. Dynamic prediction using flexible joint models of longitudinal and time-to-event data. *Stat Med*. 2017;36:1447-1460.
7. Andrinopoulou E, Eilers PH, Takkenberg JJ, Rizopoulos D. Improved dynamic prediction from joint models of longitudinal and survival data with time-varying effects using P-splines. *Biometrics*. 2018;74:685-693.
8. Hashemi R, Jacqmin-Gadda H, Commenges D. A latent process model for joint modeling of events and marker. *Lifetime Data Anal*. 2003;9:331-343.
9. Jacqmin-Gadda H, Commenges D, Dartigues JF. Random change point model for joint modeling of cognitive decline and dementia. *Biometrics*. 2006;62:254-260.
10. Li K, Luo S. Functional joint model for longitudinal and time-to-event data: an application to Alzheimer's disease. *Stat Med*. 2017;36:3560-3572.
11. Hastie T, Tibshirani R, Friedman J. *The Elements of Statistical Learning: Data Mining, Inference and Prediction*. 2nd ed. New York, NY: Springer; 2009.
12. Guo R, Zhu H, Chow S, Ibrahim JG. Bayesian lasso for semiparametric structural equation models. *Biometrics*. 2012;68:567-577.
13. Kang K, Song XY, Hu XJ, Zhu HT. Bayesian hidden Markov models for delineating the pathology of Alzheimer's disease. *Stat Med*. 2019;38:1634-1650.
14. Lang S, Brezger A. Bayesian P-splines. *J Comput Graph Stat*. 2004;13:183-212.
15. Song XY, Lu ZH. Semiparametric latent variable models with Bayesian P-splines. *J Comput Graph Stat*. 2010;19:590-608.
16. Lee EJ, Zhu HT, Kong DH, Wang YL, Giovanello KS, Ibrahim JG. BFLCRM: A Bayesian functional linear Cox regression model for predicting time to conversion to Alzheimer's disease. *Ann Appl Stat*. 2015;9:2153-2178.
17. Kong DH, Ibrahim JG, Lee EJ, Zhu HT. FLCRM: functional linear Cox regression model. *Biometrics*. 2018;74:109-117.
18. Sammel M, Ryan L. Latent variable models with fixed effects. *Biometrics*. 1996;52:650-663.
19. Roy J, Lin XH. Latent variable models for longitudinal data with multiple continuous outcomes. *Biometrics*. 2000;56:1047-1054.
20. Pan D, He HJ, Song XY, Sun LQ. Regression analysis of additive hazards model with latent variables. *J Am Stat Assoc*. 2015;110:1148-1159.
21. Pan D, Kang K, Wang CJ, Song XY. Bayesian proportional hazards model with latent variables. *Stat Methods Med Res*. 2019;28:986-1002.
22. He HJ, Pan D, Song XY, Sun LQ. Additive mean residual life model with latent variables under right censoring. *Stat Sin*. 2019;29:47-66.
23. Song XY, Lee SY. *Basic and Advanced Bayesian Structural Equation Modeling: With Applications in the Medical and Behavioral Sciences*. London, UK: Wiley; 2012.
24. Ibrahim JG, Chen MH, Sinha D. *Bayesian Survival Analysis*. New York, NY: Springer-Verlag; 2001.
25. Yin G, Ibrahim JG. A general class of Bayesian survival models with zero and nonzero cure fractions. *Biometrics*. 2005;61:403-412.
26. Pan D, Wei YY, Song XY. Joint analysis of mixed types of outcomes with latent variables. *Stat Med*. 2021;40:1272-1284.
27. Rosenberg PS. Hazard function estimation using B-splines. *Biometrics*. 1995;51:874-887.
28. Younes N, Lachin J. Link-based models for survival data with interval and continuous time censoring. *Biometrics*. 1997;53:1199-1211.
29. Zhou XX, Song XY. Mediation analysis for mixture Cox proportional hazards cure models. *Stat Methods Med Res*. 2021;30:1554-1572.
30. Cox DR, Hinkley DV. *Theoretical Statistics*. London, UK: Chapman & Hall; 1974.
31. Heagerty PJ, Zheng YY. Survival model predictive accuracy and ROC curves. *Biometrics*. 2005;61:92-105.
32. Rizopoulos D. Dynamic prediction and prospective accuracy in joint models for longitudinal and time-to-event data. *Biometrics*. 2011;67:819-829.
33. Li TF, Xie FC, Feng XN, Ibrahim JG, Zhu HT. Functional linear regression models for nonignorable missing scalar responses. *Stat Sin*. 2018;28:1867-1886.
34. Mohs RC, Knopman D, Petersen RC, et al. Development of cognitive instruments for use in clinical trials of antidementia drugs: additions to the Alzheimer's disease assessment scale that broaden its scope. *Alzheimer Dis Assoc Disord*. 1997;11:S13-S21.
35. Okuizumi K, Onodera O, Tanaka H, et al. ApoE-ε4 and early-onset Alzheimer's. *Nat Genet*. 1994;7:10-11.
36. Jöreskog KG, Sörbom D. LISREL 8: Structural Equation Modeling with the SIMPLIS Command Language. *Scientific Software International*; 1996.
37. Helmer C, Damon D, Letenneur L, et al. Marital status and risk of Alzheimer's disease: a French population-based cohort study. *Neurology*. 1999;53:1953-1958.
38. Corder EH, Saunders AM, Strittmatter WJ, et al. Gene dose of apolipoprotein E type 4 allele and the risk of Alzheimer's disease in late onset families. *Science*. 1993;261:921-923.
39. Saunders AM, Strittmatter WJ, Schmechel D, et al. Association of apolipoprotein E allele ε4 with late-onset familial and sporadic Alzheimer's disease. *Neurology*. 1993;43:1467-1472.
40. Podcasy JL, Epperson CL. Considering sex and gender in Alzheimer disease and other dementias. *Dialogues Clin Neurosci*. 2016;18:437-446.
41. Viña J, Lloret A. Why women have more Alzheimer's disease than men: gender and mitochondrial toxicity of amyloid-β peptide. *J Alzheimers Dis*. 2010;20:S527-S533.
42. Chen P, Ratcliff G, Belle SH, Cauley JA, Dekosky ST, Ganguli M. Patterns of cognitive decline in presymptomatic Alzheimer disease: a prospective community study. *Arch Gen Psychiatry*. 2001;58:853-858.
43. Cloutier S, Chertkow H, Kergoat M, Gauthier S, Belleville S. Patterns of cognitive decline prior to dementia in persons with mild cognitive impairment. *J Alzheimers Dis*. 2015;47:901-913.

44. Birks J. Cholinesterase inhibitors for Alzheimer's disease. *Cochrane Database Syst Rev*. 2006;CD005593. 1.
45. Neal M, Barton WP. Validation therapy for dementia. *Cochrane Database Syst Rev*. 2003;3:CD001394.
46. Kutner MH, Nachtsheim CJ, Neter J. *Applied Linear Regression*. 4th ed. New York, NY: McGraw-Hill/Irwin Series; 2004.
47. Weiss NA. *Introductory Statistics*. 10th ed. Boston, MA: Pearson; 2012.
48. Escobar M. Estimating normal means with a Dirichlet process prior. *J Am Stat Assoc*. 1994;89:268-277.
49. Grambsch PM, Therneau TM. Proportional hazards tests and diagnostics based on weighted residuals. *Biometrika*. 1994;81:515-526.

SUPPORTING INFORMATION

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

How to cite this article: Kang K, Pan D, Song X. A joint model for multivariate longitudinal and survival data to discover the conversion to Alzheimer's disease. *Statistics in Medicine*. 2021;1-18. doi: 10.1002/sim.9241