Received: 18 March 2021 | Revised: 27 February 2022 | Accepted: 26 April 2022

DOI: 10.1002/sim.9433

RESEARCH ARTICLE

Statistics in Medicine WILEY

Dynamic prediction with time-dependent marker in survival analysis using supervised functional principal component analysis

Haolun Shi[1] | **Shu Jiang**[2] | **Jiguo Cao**[1]

[1]Department of Statistics and Actuarial Science, Simon Fraser University, Burnaby, British Columbia, Canada

[2]Division of Public Health Sciences, Washington University School of Medicine in St. Louis, St. Louis, Missouri, USA

Correspondence
Jiguo Cao, Department of Statistics and Actuarial Science, Simon Fraser University, Burnaby, BC, Canada.
Email: jiguo_cao@sfu.ca

Funding information
Canadian Network for Research and Innovation in Machining Technology, Natural Sciences and Engineering Research Council of Canada, Grant/Award Number: RGPIN-2018-06008; Division of Cancer Prevention, National Cancer Institute, Grant/Award Number: R37 CA256810

Time-varying biomarkers reflect important information on disease progression over time. Dynamic prediction for event occurrence on a real-time basis, utilizing time-varying information, is crucial in making accurate clinical decisions. Functional principal component analysis (FPCA) has been widely adopted in the literature for extracting features from time-varying biomarker trajectories. However, feature extraction via FPCA is conducted independent of the time-to-event response, which may not produce optimal results when the goal lies in prediction. With this consideration, we propose a novel supervised FPCA, where the functional principal components are determined to optimize the association between the time-varying biomarker and time-to-event outcome. The proposed framework also accommodates irregularly spaced and sparse longitudinal data. Our method is empirically shown to retain better discrimination and calibration performance than the unsupervised FPCA method in simulation studies. Application of the proposed method is also illustrated in the Alzheimer's Disease Neuroimaging Initiative database.

KEYWORDS
dynamic prediction, functional principal component analysis, supervised learning, time-varying covariates

1 | INTRODUCTION

Precision medicine is an emerging field for the future health care where the goal lies in clinical decision making that accommodates individual-specific variability. The intrinsic heterogeneity between and within patients over time are reflected in part, by the time-varying covariates, which provide important information for risk prediction. These covariates can be viewed as functions of time, which are also called functional covariates.[1,2] Making use of both the demographic and functional covariate is thus essential in making more accurate clinical decisions on the individual level. The challenge, however, lies in modeling the functional covariate, which is theoretically infinite dimensional. Thus, finding the most "informative" low-dimensional features is an essential task in practice.

Functional principal component analysis (FPCA) has been a popular dimension reduction technique. FPCA identifies the major mode of variations among functional data, which are represented by the functional principal components (FPCs). The infinite-dimensional functional data are projected into a low-dimensional space spanned by the FPCs.[1,2] Ramsay and Dalzell[3] and Rice and Silverman[4] developed smoothed FPCA methods to control the smoothness of FPCs. Yao et al[5] proposed to calculate the FPC scores based on conditional expectation when functional data are collected on

sparse grid. Chen and Lei,[6] Lin et al,[7] and Nie and Cao[8] proposed to estimate FPCs which are only nonzero in a small interval in order to enhance the interpretability of FPCs. Other studies involving FPCA jointly with survival data can be found in Yan et al,[9] Wang et al,[10] and Jiang et al,[11] for example. However, FPCA does not consider the relationship between the functional predictor and the response variable. Thus, the ordering of the FPCs do not indicate the degree of association with the outcome measure. With this consideration, it warrants further research to explore supervised methods, where it will likely improve the model prediction performance.

More recently, Li et al[12] proposed the supervised sparse functional principal component analysis. They assumed that additional scalar (continuous) supervision variables drive low-rank structure of the functional predictor. Nie et al[13] considered supervised FPCA with the presence of the scalar response variable (continuous and binary). They were interested in identifying set of FPCs to boost the prediction performance. To the best of our knowledge, no method has been proposed to accommodate time-to-event data with sparse longitudinal observations.

Extending the FPCA method to be supervised by survival outcome is not trivial due to the presence of right-censoring, irregularly spaced and often sparse longitudinal data. We are thus motivated to propose a novel supervised FPCA framework in this article by constructing an objective function that incorporates the association between functional predictor and time-to-event outcome utilizing inverse probability censoring weights. The FPC scores are framed as conditional expectations to accommodate potential sparsity.[5] An optimization procedure that utilizes eigenvalue decomposition is provided in this article that warrants great computational efficiency. The proposed method is demonstrated to be versatile in simulation studies, where effective dimension reduction is achieved by choosing the top few supervised functional principal components that cumulatively explain a large proportion of association with the failure time.

The rest of this article is organized as follows. We introduce notation and model setup under the supervised FPCA framework incorporating time-to-event outcomes in Section 2. Specifically, we give detailed discussion on the implementation of proposed algorithm via eigenvalue decomposition, and draw connection between supervised FPCA and dynamic risk prediction. In Section 3, we conduct intensive simulation studies to investigate the finite sample performance of the proposed method. The utility of the proposed method is illustrated with data from the Alzheimer's Disease Neuroimaging Initiative (ADNI) in Section 4. We conclude with discussion in Section 5.

2 | NOTATION AND METHOD

Let T_i and C_i denote the time of event and censoring for an individual i, respectively. We denote the observed time by $\tilde{T}_i = \min(T_i, C_i)$, and let $\triangle_i = I(T_i < C_i)$ be the indicator of event occurrence. We further let X_i denote a vector of fixed covariates of dimension $P \times 1$. We consider a single time-varying covariate Z_i in this article, where $Z_i = \{Z_i(t_{i1}), \ldots, Z_i(t_{i,R_i})\}$ represents the observed values at R_i distinct occasions for each individual i, $i = 1, \ldots, n$.

2.1 | Supervised FPCA with time-to-event data

Under the functional framework, we assume that the discrete observations Z_i are realizations of a stochastic process $Z(t)$ in a square integrable functional space with domain $t \in [0, \tau]$. It is well known that this is a separable Hilbert space with the inner product defined as $\langle f, g\rangle = \int_0^\tau f(t)g(t)dt$ for $f, g \in L^2([0, \tau], \text{I R})$. In general, the stochastic process is assumed to have mean function $E[Z(t)] = \mu(t)$ and covariance operator $C(t, s) = \text{cov}(Z(t), Z(s))$ for $\forall t, s \in [0, \tau]$. Then by Mercer's theorem,[14]

$$C(t, s) = \sum_k^\infty v_k \phi_k(s)\phi_k(t) \quad \forall t, s \in [0, \tau]\,, \tag{1}$$

where $\phi_k(t)$ is the kth orthonormal basis function and v_k is the corresponding eigenvalue, $v_1 \geq v_2, \geq \cdots > 0, k = 1, \ldots \infty$. Without loss of generality, we assume that the process is de-meaned, that is, $E[Z(t)] = 0$, from here on. Then by Karhunen-Loève theorem,[15] each curve $Z_i(t)$ can be characterized by,

$$Z_i(t) = \sum_k^\infty \xi_{i,k}\phi_k(t), \quad \forall t \in [0, \tau],$$

where $\xi_{i,k} = \langle Z_i, \phi_k\rangle$ is the kth score for individual i.

Under the supervised FPCA framework, we aim to allocate a specific set of orthonormal basis functions $\phi = (\phi_1, \phi_2, \dots)$, that is, $||\phi|| = 1$, $\langle \phi_k, \phi_{k'} \rangle = 0$ for $k < k'$, such that $Q(\phi)$ is maximized,

$$Q(\phi) = \frac{\theta \text{var}\langle Z, \phi \rangle + (1-\theta)\text{cov}(\log(\tilde{T}), \langle Z, \phi \rangle)}{||\phi||^2}, \tag{2}$$

for $0 < \theta \leq 1$. Note that $\theta = 1$ forces $Q(\phi)$ to be the objective function for the conventional FPCA; whereas when $\theta \neq 1$, the objective function accommodates covariance between $\log(\tilde{T}_i)$ and the functional scores. The squared covariance term ensures that the objective function is always positive and facilitates eigenvalue decomposition as demonstrated in the next subsection. The optimal value of θ can be determined from tuning the objective function via cross-validation. In practice, it is not feasible to work with an infinite sequence of basis functions, thus a suitable truncation to the first $K^\star$ terms is needed. To choose the suitable point of truncation, we may first plot the objective function $Q(\phi)$ against $K^\star$ and then locate the "elbow point" where the function flattens out as our threshold for truncation.

The conventional covariance estimator between two random variables is not appropriate in (2), as this would result in biased estimators in the presence of right censoring. Instead, the inverse probability censoring weight (IPCW) needs to be applied onto the ith observation as,

$$w_i = \frac{\Delta_i}{\hat{G}(\log(\tilde{T}_i))},$$

where $\hat{G}(t) = P(\log(C) > t)$, $i = 1, \dots, n$.[16] The Kaplan-Meier estimator can be used as a consistent estimator of the censoring probabilities under the assumption of independent censoring. The covariance term can then be estimated by,

$$\text{cov}(\log(\tilde{T}), \langle Z, \phi \rangle) = E_i \left\{ w_i \langle Z_i, \phi \rangle (\log(\tilde{T}_i) - \overline{T}) \right\},$$

where $\overline{T} = E_i\{w_i \log(\tilde{T}_i)\}$. This is known to be consistent estimator of the covariance.[17]

2.2 | Estimation of supervised FPCA

Consider n individuals each with observed data $\mathcal{D}_i = \{(\tilde{T}_i, \Delta_i), X_i, Z_i, i = 1, \dots, n$. For each individual i, the longitudinal predictor $Z_i = \{Z_i(t_{i1}), \dots, Z_i(t_{i,R_i})\}$ is measured at some finite number of observation times $t_i = (t_{i1}, \dots, t_{i,R_i})$. We can define a finite grid as $\bigcup_{i=1}^n t_i$, covering all unique observation times for the sample, with $\tau = \max(\bigcup_{i=1}^n t_i)$. Assuming we have the de-meaned samples $Z_1, \dots, Z_n$, the proposed optimization procedure can be stated in five steps given a pre-specified value of θ.

1. Start with an arbitrary set of smooth basis functions $\psi_1(t), \dots, \psi_K(t)$ to characterize the functional predictor, where K denotes the total number of basis functions. The set of basis functions can be, for example, the empirical basis functions estimated by the conventional FPCA method, but can be flexibly extended to other basis functions. The functional data can then be rewritten in matrix notation as $(Z_1(t), \dots, Z_n(t))^T = \lambda\psi(t)$, where $\lambda = (\lambda_1, \dots, \lambda_n)^T$, with $\lambda_i = (\lambda_{i,1}, \dots, \lambda_{i,K})^T$, and $\psi(t) = (\psi_1(t), \dots, \psi_K(t))^T$, $t \in [0, \tau]$.
2. Rewrite the supervised basis function as $\phi_k(t) = \gamma_k^T \psi(t)$, where $\gamma_k = (\gamma_{k,1}, \dots, \gamma_{k,K})^T$ is the coefficient vector. Here, we assume that the set of supervised basis function can be found by linear combinations of $\psi(t)$. The set of scores in (2) can then be written as,

$$\langle Z_i, \phi \rangle = \boldsymbol{\gamma}^T \boldsymbol{M} \lambda_i = \boldsymbol{\gamma}^T \boldsymbol{M}_i, \tag{3}$$

   where $\boldsymbol{\gamma} = (\gamma_1, \dots, \gamma_K)^T$, λ_i is the ith row of λ, and $\boldsymbol{M}$ is of dimension $K \times K$, with the (k, k') entry being $\langle \psi_k(t), \psi_{k'}(t) \rangle$ for $k, k' \in K$, $t \in [0, \tau]$.
3. The variance of (3) can then be estimated empirically as,

$$\text{var}(\langle Z, \phi \rangle) = \frac{1}{n} \boldsymbol{\gamma}^T \boldsymbol{M} \lambda^T \lambda \boldsymbol{M} \boldsymbol{\gamma}. \tag{4}$$

The covariance term can be represented in a similar fashion,

$$\text{cov}(Y, \langle Z, \phi \rangle) = \frac{1}{n} \boldsymbol{\gamma}^T \boldsymbol{M} \boldsymbol{\lambda}^T (Y \circ w), \tag{5}$$

where $Y = ((\log(\widetilde{T_1}) - \overline{T}), \ldots, (\log(\widetilde{T_n}) - \overline{T}))$ is the vector of de-meaned outcome measures, with $w = (w_1, \ldots, w_n)^T$ denoting the vector of IPCW weights, and $(a \circ b)$ the element-wise product between vectors a and b.

4. Put (4) and (5) together to reconstruct the objective function in (2),

$$Q(\phi) = \frac{\boldsymbol{\gamma}^T \boldsymbol{U} \boldsymbol{\gamma}}{\boldsymbol{\gamma}^T \boldsymbol{M} \boldsymbol{\gamma}}, \tag{6}$$

where

$$\boldsymbol{U} = \frac{\theta}{n} \boldsymbol{M} \boldsymbol{\lambda}^T \boldsymbol{\lambda} \boldsymbol{M} + \frac{1-\theta}{n^2} \boldsymbol{M} \boldsymbol{\lambda}^T (Y \circ w)(Y \circ w)^T \boldsymbol{\lambda} \boldsymbol{M}^T .$$

Note that maximizing (6) is equivalent to maximizing,

$$\boldsymbol{\delta}^T (\boldsymbol{M}^{-1/2})^T \boldsymbol{U} \boldsymbol{M}^{-1/2} \boldsymbol{\delta},$$

subject to $\boldsymbol{\delta}^T \boldsymbol{\delta} = 1$, where $\delta = \boldsymbol{M}^{1/2} \boldsymbol{\gamma}$.

5. Estimate $\boldsymbol{\delta}$ by finding the leading $\delta_1, \ldots, \delta_{K^\star}$ eigenvectors of the matrix $(\boldsymbol{M}^{-1/2})^T \boldsymbol{U} \boldsymbol{M}^{-1/2}$. As a result, we are able to estimate $\hat{\gamma}_k = \boldsymbol{M}^{-1/2} \delta_k$, and consequently the supervised basis functions,

$$\hat{\phi}_k(t) = \hat{\gamma}_k^T \psi(t) ,$$

for $k = 1, \ldots, K^\star$, $t \in [0, \tau]$. The truncation $K^\star$ can be chosen from an elbow plot such that increasing $K^\star$ will not significantly increase the objective function value in (6).

2.3 | Survival analysis incorporating the sparse longitudinal covariate

Given the set of supervised basis functions $\hat{\phi} = (\hat{\phi}_1, \ldots, \hat{\phi}_{K^\star})^T$, we can construct a survival model using the demographic predictors X_i and the set of individual-specific functional scores $\hat{\xi}_i = (\hat{\xi}_{i,1}, \ldots, \hat{\xi}_{i,K^\star})^T$ as fixed covariates, $i = 1, \ldots, n$. When the observed functional predictor is sufficiently dense on an regularly spaced time grid, the scores can be easily obtained from $\hat{\xi}_{i,k} = \langle \hat{\phi}_k, Z_i \rangle$. With sparse and irregularly observed longitudinal data, the scores can be approximated via sums as $\hat{\xi}_{ik} = \sum_{j=1}^{R_i} Z_i(t_{ij}) \hat{\phi}_k(t_{ij})(t_{ij} - t_{i,j-1})$. As proposed in Yao et al,[5] such summation may not provide reasonable approximations when individuals has only two observations and when there is measurement error present. Thus, under the Gaussian assumption,[18] we can estimate the best linear unbiased predictor (BLUP) for the scores as,

$$\begin{aligned} \hat{\xi}_{ik} &= E[\xi_{ik} | Z_i] \\ &= E[\xi_{ik}] + \text{cov}(\xi_{ik}, Z_i) \text{cov}(Z_i, Z_i)^{-1} Z_i \\ &= \hat{v}_k \hat{\phi}_{ik}^T \widehat{\text{cov}}(Z_i, Z_i)^{-1} Z_i , \end{aligned}$$

where $\hat{v}_k$ is the kth leading eigenvalue of the matrix $(\boldsymbol{M}^{-1/2})^T \boldsymbol{U} \boldsymbol{M}^{-1/2}$, and $\hat{\phi}_{ik} = (\hat{\phi}_k(t_{i1}), \ldots, \hat{\phi}_k(t_{i,R_i}))^T$, $k = 1, \ldots, K^\star$.

For simplicity, we present a Cox proportional hazards model in this article, but the model choice can be flexible and adapt to the data. The survival distribution at any time t can be written as,

$$S_0(t)^{\exp(\alpha' X_i + \beta' \hat{\xi}_i)} ,$$

where $S_0(t) = \exp(-H_0(t))$, with $H_0(t)$ being the cumulative baseline hazard function, $\alpha = (\alpha_1, \ldots, \alpha_P)^T$, and $\beta = (\beta_1, \ldots, \beta_{K^\star})^T$ are the regression coefficients. The cumulative baseline hazard function under the Cox model can be estimated using the Breslow estimator,[19,20]

$$\widehat{H}_0(t) = \sum_{T_{(l)} \le t} \frac{d_{(l)}}{\sum_{i \in \mathcal{R}(T_{(l)})} \exp\left(\alpha' X_i + \beta' \hat{\xi}_i\right)},$$

where $T_{(1)} < T_{(2)} < \cdots$ denote the ordered distinct event times, $d_{(l)}$ denote the number of events at $T_{(l)}$, and $\mathcal{R}(T_{(l)})$ is the risk set consisting of all the individuals still susceptible to the event at time $T_{(l)}$.

To make individualized dynamic predictions, we assume that there are n individuals in the training cohort where we first obtain the estimated $\hat{\alpha}$, $\hat{\beta}$, and $\hat{H}_0(t)$. We then consider an $(n+1)$th individual who is event-free with observed data X_{n+1} and longitudinal observations up to some time $t^\star$, $t^\star < \tau$. At any time $t^\star + \Delta t$, $t^\star + \Delta t < \tau$, the predicted future survival distribution conditional on the observed data can be written as,

$$P(T_{n+1} \ge t^\star + \Delta \mid T_{n+1} > t^\star, X_{n+1}, \hat{\xi}_{n+1}) = \left\{ \frac{\hat{S}_0(t^\star + \Delta t)}{\hat{S}_0(t^\star)} \right\}^{\exp(\hat{\alpha}' X_{n+1} + \hat{\beta}' \hat{\xi}_{n+1})},$$

where the $(n+1)$th individual's functional score $\hat{\xi}_{n+1}$ can be estimated by

$$\hat{\xi}_{n+1,k} = \hat{\nu}_k \hat{\phi}_{n+1,k}^T \widehat{\text{cov}}(Z_{n+1}, Z_{n+1})^{-1} Z_{n+1},$$

with $\hat{\nu}_k$ estimated from training data, and $\hat{\phi}_{n+1,k}$, $\widehat{\text{cov}}(Z_{n+1}, Z_{n+1})$ obtained from projecting individual $(n+1)$'s data onto the trained $\hat{\phi}$ and covariance function.

3 | SIMULATION STUDY

3.1 | Simulation setup

We aim to investigate the model predictive performance under the proposed supervise FPCA (sFPCA) in comparison with both joint model (JM) and the conventional FPCA in this simulation study.[5,21] We first define the set of true basis functions,

$$\psi_1(t) = \sqrt{2}\cos(2\pi t),$$
$$\psi_2(t) = \sqrt{2}\sin(4\pi t),$$
$$\psi_3(t) = \sqrt{2}\cos(4\pi t),$$

such that the constraints $||\psi_k||^2 = 1$ and $\langle \psi_k, \psi_{k'} \rangle = 1$ if $k = k'$, and 0 otherwise, are fulfilled, $k = 1, 2, 3$. We then independently sample the scores according to $\lambda_i \sim MVN(0, \Sigma)$, where $\Sigma = \text{diag}(10, 6, 3)$. Given the set of true basis functions and scores, the longitudinal trajectory can be formulated according to the Karhunen-Loève expansion as,

$$Z_i(t) = \mu(t) + \lambda_{i,1}\psi_1(t) + \lambda_{i,2}\psi_2(t) + \lambda_{i,3}\psi_3(t),$$

where the mean function $\mu(t)$ is assumed to be 0. The individualized realizations of the longitudinal trajectory $\{Z_i(t_{i,r}), r = 1, \ldots, R_i\}$ are assumed to have $\max(R_i) \le 20$ for $\forall i$, constrained by censoring or event occurrence. Similar to the application setting, we consider these R_i visits to happen on a fixed time grid from 0 to 25 with an increment of $25/\max(R_i)$ unit.

To link covariates to the time-to-event outcome, we assume a proportional hazards model such that the hazard function follows,

$$h_i(t) = h_0(t) \exp\left\{ \alpha_1 X_i + \int_0^\tau \phi(t) Z_i(t) dt \right\},$$

where τ is the maximum observation time. The fixed covariate X_i is assumed to follow a Bernoulli distribution with success probability of 0.5, with the corresponding coefficient α_1 set to -1. We consider two potential scenarios for the time-varying coefficient:

$$\text{Scenario 1}: \ \phi(t) = 0.25\psi_1(t) + 0.5\psi_2(t) + \psi_3(t)\,,$$
$$\text{Scenario 2}: \ \phi(t) = 0\psi_1(t) + 0\psi_2(t) + \psi_3(t)\,.$$

Under scenario 1, the hazard function is influenced by a linear combination of all three basis functions; whereas we consider a more extreme case by constraining the hazard function to be influenced only by the third basis function $\psi_3(t)$ under scenario 2. Here, we let the baseline hazard follow a Weibull distribution $h_0(t) = \kappa\rho(\rho t)^{\kappa-1}$ with increasing risk over time, and consider $\kappa = 2$, $\rho = 0.096, 0.075$ for scenarios 1 and 2, respectively, in order to satisfy the censoring constraint as we discuss below. Given the above set-up, the survival time T_i can then be generated from the inverse of the cumulative hazard function $H_i^{-1}(u)$, where $u \sim \text{unif}(0, 1)$. We have assumed the independent censoring scheme in this simulation study, where $C_i \sim \text{unif}(0, C_{\max})$, with $C_{\max}$ set at a value such that the % of being censored by the end of the study approximately matches our target censoring percentage.

As for the scheme of missingness, we consider both missing completely at random (MCAR) and missing at random (MAR). If the probability of being missing is the same for all individuals, then the data are said to be missing completely at random. If the probability of being missing is the same only within a certain group defined by the observed data, then the data are said to be missing at random. For MCAR scheme, we randomly choose the data points and assign them to be missing, and approximately 10% of the longitudinal observations are set to be missing. For MAR scheme, we assign different probabilities of being missing based on value of the simulated binary variable X_i: if $X_i = 0$, then the missing probability is 20%, otherwise the missing probability is 10%.

3.2 | Prediction accuracy measures

To assess and compare the model prediction performance, we have adopted the widely used prediction accuracy measure, time-dependent AUC and Brier score (BS), in our simulation study.[22] The time-dependent AUC serves as a discrimination measure of how well the predicted risk distinguishes between subjects who are more likely to experience the event in time window (s, t) than the subjects who are less likely, conditional on the survival up to time s. The time-dependent AUC at landmark time s for a prediction time t is thus defined as

$$AUC(s, t) = P(\pi_i(s, t) > \pi_j(s, t) \mid T_i < t, T_j > t, T_i > s, T_j > s),$$

where $\pi_i(s, t)$ is the predicted probability of event at time t conditional on the observed data sand survival up to time s. The Brier score is a conditional mean squared error of the predicted risk vs the actual event given survival beyond time s, and its definition is given by

$$BS(s, t) = E\left[\{1(T_i > t) - \pi_i(s, t)\}^2 \mid T_i > s\right].$$

For our case, we take $s = t^\star$ and $t = t^\star + 5$. As the failure time T_i is not always observable due to censoring, the inverse probability of censoring weighting techniques are commonly used to obtain consistent estimators of the AUC and BS. For a comprehensive elaboration of the details of calculation of these two predictive measures, the interested reader are referred to Schoop et al,[23] Hung and Chiang,[24] and Blanche et al.[25]

3.3 | Simulation results

To avoid over-fitting, we have randomly chosen 33% of the individuals from each dataset to be the testing set, and the rest as the training set. We consider sample sizes of $n = 300$ and $n = 600$ individuals. The predictive accuracy metrics are averaged based on 500 data replications. We consider cases where the censoring percentage approximately equals 30% or 50%. For FPCA and sFPCA, we consider both cases of $K^\star = 1$ and 2, that is, using only the first 1 to 2 FPC scores for building predictive models. For a fair comparison, we have chosen the same number of FPCs (1 or 2) for both the supervised and conventional FPCA. As for the scheme of missingness, both the MCAR and MAR are considered. We compare the prediction performance of sFPCA with that of the JM approach and the conventional FPCA.

Within the training set, the tuning parameter θ was chosen from an equal distanced grid in $[0, 1]$ with an increment of 0.05 under the five-fold cross validation, where the optimal value was determined to be the one that gives the best prediction accuracy. The testing set is then used to record the dynamic prediction performance for each time window $(t^\star, t^\star + 5]$ conditional on data observed up to $t^\star$, $t^\star = 6, 8, 10$. Note that individuals in the testing set need to have their time-varying trajectories beyond $t^\star$ removed, prior to projecting back onto the trained parameters in estimating their scores, to mimic the real data setting.

Tables 1 and 2 present the mean AUC and BS by selecting the first $K^\star = 1, 2$ basis functions, conditional on data observed prior to $t^\star = 6, 8, 10$ in forecasting $t^\star + 5$ under a sliding window framework, when the longitudinal data are MCAR and MAR, respectively. The true AUC is estimated using the full trajectory for training without any missingness and shortened trajectory due to censoring. We also record the mean number of visits in each longitudinal trajectory prior to $t^\star$ in the testing set, as well as the mean of the selected tuning parameter θ. The overall total numbers of visits (including the ones beyond time $t^\star$) are 18 and 17 under the MCAR and MAR schemes, respectively. Overall, we observe that in almost all cases, the supervised FPCA achieves a higher AUC and a lower BS than the other two models, indicating its advantage in predictive accuracy. The computing speed is very promising for the proposed method. For one simulation

TABLE 1 Estimated mean AUC$(t^\star, t^\star + \Delta t)$ and BS$(t^\star, t^\star + \Delta t)$ using the joint model (JM), conventional FPCA, and supervised FPCA by selecting the first $K^\star$ basis functions, conditional on data observed prior to $t^\star = 6, 8, 10$ in forecasting $t^\star + 5$ under a sliding window framework when the longitudinal data are missing completely at random (MCAR) with training/testing sample size of 400/200; visits denote the mean number of observations prior to $t^\star$ in the testing set, C denotes the censoring percentage; the mean of the selected tuning parameter θ are shown as well

Scen	C	$t^\star$	$K^\star$	visits	AUC_{true}	AUC_{JM}	AUC_{FPCA}	$\text{AUC}_{\text{sFPCA}}$	BS_{JM}	BS_{FPCA}	BS_{sFPCA}	θ
1	0.3	6	1	4.66	0.880	0.606	0.662	0.861	0.179	0.208	0.157	0.980
			2				0.783	0.862		0.180	0.154	0.987
		8	1	6.44	0.870	0.598	0.592	0.824	0.166	0.200	0.160	0.970
			2				0.695	0.838		0.187	0.154	0.991
		10	1	7.33	0.864	0.597	0.564	0.788	0.164	0.191	0.161	0.992
			2				0.679	0.829		0.180	0.152	0.992
	0.5	6	1	4.66	0.882	0.621	0.717	0.865	0.149	0.171	0.135	0.937
			2				0.790	0.866		0.154	0.130	0.907
		8	1	6.44	0.869	0.610	0.623	0.821	0.137	0.169	0.137	0.964
			2				0.720	0.846		0.155	0.128	0.843
		10	1	7.33	0.873	0.608	0.589	0.788	0.131	0.161	0.136	0.954
			2				0.699	0.838		0.147	0.124	0.838
2	0.3	6	1	4.66	0.860	0.538	0.600	0.694	0.236	0.240	0.221	0.906
			2				0.634	0.704		0.233	0.219	0.940
		8	1	6.44	0.846	0.536	0.535	0.658	0.229	0.236	0.220	0.861
			2				0.583	0.704		0.233	0.213	0.921
		10	1	7.33	0.835	0.537	0.499	0.622	0.230	0.229	0.218	0.819
			2				0.566	0.704		0.226	0.205	0.931
	0.5	6	1	4.66	0.856	0.551	0.604	0.689	0.207	0.208	0.193	0.709
			2				0.644	0.716		0.202	0.188	0.792
		8	1	6.45	0.844	0.551	0.546	0.652	0.196	0.204	0.190	0.657
			2				0.610	0.737		0.196	0.176	0.751
		10	1	7.34	0.829	0.552	0.516	0.619	0.193	0.194	0.184	0.588
			2				0.590	0.732		0.188	0.167	0.717

TABLE 2 Estimated mean AUC($t^\star, t^\star + \triangle t$) and BS($t^\star, t^\star + \triangle t$) using the joint model (JM), conventional FPCA, and supervised FPCA by selecting the first $K^\star$ basis functions, conditional on data observed prior to $t^\star = 6, 8, 10$ in forecasting $t^\star + 5$ under a sliding window framework when the longitudinal data are missing at random (MAR) with training/testing sample size of 400/200; visits denote the mean number of observations prior to $t^\star$ in the testing set, C denotes the censoring percentage; the mean of the selected tuning parameter θ are shown as well

Scen	C	$t^\star$	$K^\star$	visits	AUC_{true}	AUC_{JM}	AUC_{FPCA}	$\text{AUC}_{\text{sFPCA}}$	BS_{JM}	BS_{FPCA}	BS_{sFPCA}	θ
1	0.3	6	1	4.52	0.880	0.606	0.664	0.862	0.180	0.209	0.157	0.990
			2				0.790	0.863		0.179	0.154	0.973
		8	1	6.20	0.870	0.598	0.583	0.826	0.166	0.200	0.159	0.981
			2				0.700	0.838		0.185	0.155	0.975
		10	1	7.05	0.864	0.597	0.561	0.790	0.163	0.191	0.161	0.981
			2				0.684	0.829		0.179	0.153	0.976
	0.5	6	1	4.52	0.882	0.621	0.716	0.866	0.150	0.171	0.135	0.915
			2				0.798	0.866		0.152	0.130	0.932
		8	1	6.21	0.869	0.611	0.617	0.823	0.137	0.170	0.137	0.915
			2				0.729	0.844		0.154	0.127	0.913
		10	1	7.06	0.873	0.609	0.589	0.787	0.130	0.161	0.136	0.927
			2				0.704	0.841		0.147	0.123	0.894
2	0.3	6	1	4.52	0.860	0.537	0.596	0.693	0.237	0.239	0.221	0.849
			2				0.638	0.702		0.232	0.219	0.901
		8	1	6.21	0.846	0.536	0.537	0.660	0.229	0.236	0.220	0.769
			2				0.587	0.702		0.232	0.212	0.902
		10	1	7.06	0.835	0.537	0.508	0.630	0.231	0.230	0.217	0.761
			2				0.571	0.697		0.225	0.206	0.912
	0.5	6	1	4.52	0.856	0.551	0.610	0.690	0.207	0.206	0.193	0.679
			2				0.655	0.714		0.200	0.188	0.761
		8	1	6.22	0.844	0.550	0.554	0.649	0.196	0.203	0.190	0.658
			2				0.624	0.734		0.195	0.177	0.727
		10	1	7.07	0.829	0.552	0.525	0.613	0.193	0.194	0.184	0.631
			2				0.592	0.730		0.187	0.167	0.729

replication, it takes approximately 2 min to perform the supervised FPCA on a machine with Intel i5-5200 CPU and 12 Gigabytes of RAM.

Figures 1 and 2 show the box plots of AUC and BS for JM, FPCA and supervised FPCA based on two basis functions under scenarios 1 and 2, respectively, with MCAR and a 30% censoring rate. It is apparent that the supervised FPCA out-performs the conventional FPCA dynamically over all time windows when we choose 2 FPCs as shown in Figures 1 and 2. Similar conclusions can be drawn when we choose only one FPC under both scenarios; these results are included in Figures S1 and S2 in the Supplemental Material for interested readers.

In addition, we display the numerical results for the sample size of 300 in Tables S1 and S2 in the Supplemental Material, which lead to a similar conclusion that the supervised FPCA is superior in terms of predictive accuracy. Moreover, to examine the variance of $\hat{\xi}_i$ over time, we additionally reported the empirical variance of the $\hat{\xi}_i$ for $t^\star = 6, 8, 10$ in Table S3 in the Supplementary Material. We found that the estimation of the scores $\hat{\xi}_i$ appears to be more stable with a lower censoring rate. Figure S3 in the Supplementary Material shows the trajectories of the first and second FPC scores of 30 randomly chosen individuals in the testing set in one data replication, and it can be seen that the scores are relatively stable. In practice, we recommend that such a progression of the estimated FPC scores over the landmark time $t^\star$ should be examined to ensure the reliability of the prediction.

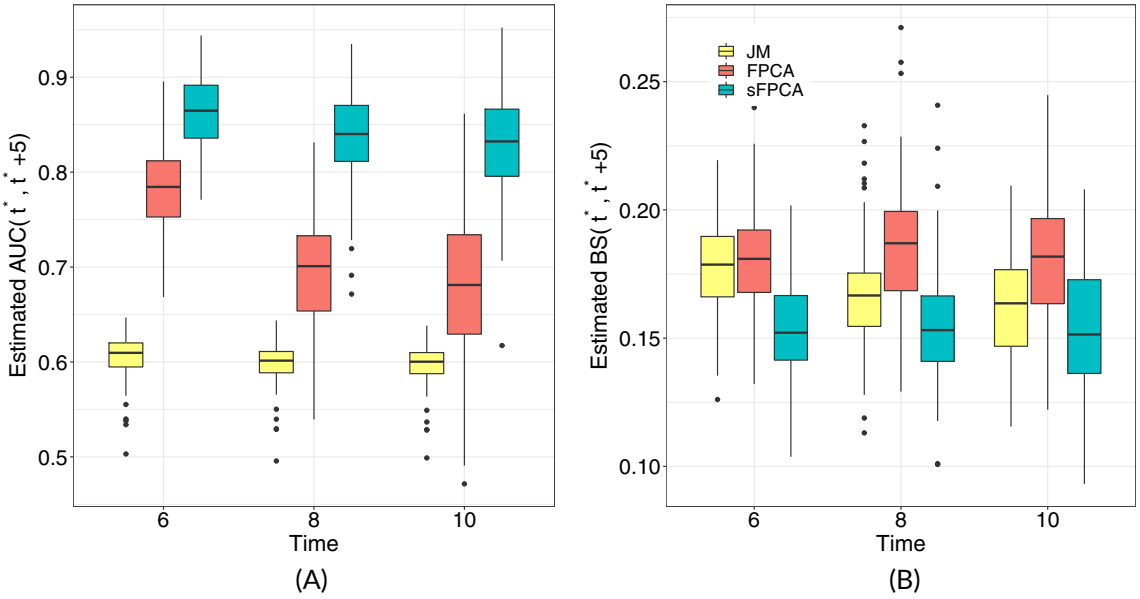

FIGURE 1 Estimated AUC($t^\star, t^\star + 5$) and BS($t^\star, t^\star + 5$) for supervised FPCA vs JM and FPCA for scenario 1 (missing completely at random (MCAR) and 30% censored), with two basis functions, conditional on data observed prior to $t^\star = 6, 8, 10$ in forecasting $t^\star + 5$ under a sliding window framework. (A) AUC; (B) Brier score

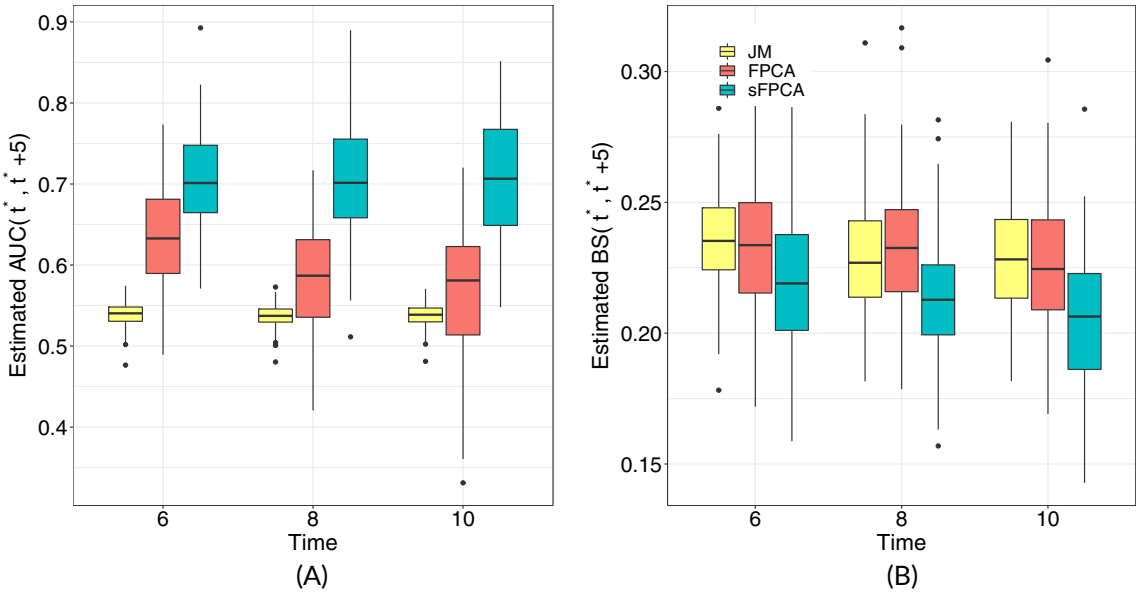

FIGURE 2 Estimated AUC($t^\star, t^\star + 5$) and BS($t^\star, t^\star + 5$) for supervised FPCA vs JM and FPCA for scenario 2 (missing completely at random (MCAR) and 30% censored), with two basis functions, conditional on data observed prior to $t^\star = 6, 8, 10$ in forecasting $t^\star + 5$ under a sliding window framework. (A) AUC; (B) Brier score

4 | ALZHEIMER'S DISEASE NEUROIMAGING INITIATIVE

The proposed method in previous sections is applied to data from the Alzheimer's Disease Neuroimaging Initiative (ADNI)*. We focus on a subset of patients who have been diagnosed with mild cognitive impairment (MCI) at the baseline, and treat the onset of dementia (AD) as the time-to-event outcome. This include a total 365 patients diagnosed with MCI at the baseline in ADNI-1, where 173 of them progressed to AD before the end of the study. Patients were assessed every 6 month since baseline with a maximum of 36 month in ADNI-1 and additional

annual follow-ups in ADNI-2, resulting in an average follow-up of 24.9 months. Correspondingly, the number of visits in this subset has a mean of 4.6 with SD of 1.5. The visit times are fixed as the multiples of 6 months. We focus on the Alzheimer's Disease Assessment Scale-Cognitive 13 items (ADAS-13) as the time-varying neurocognitive marker for each visit. The ADAS-13 scores are commonly measured in AD studies, where these scores are assumed to have a strong predictive value to patient cognition status (higher value is worse).[26] For illustration, we present in Figure S3 in the Supplemental Material, the raw longitudinal trajectories for 50 randomly selected MCI patients.

Additionally, we consider the baseline age, gender, years of education, and the presence of APOE4 alleles in our prediction model. We specify a Cox proportional hazard model as,

$$h_i(t) = h_0(t) \exp\left\{\alpha_1 Age_i + \alpha_2 Gender_i + \alpha_3 Edu_i + \alpha_4 APOE4_i + \boldsymbol{\beta}^T \hat{\xi}_i\right\} ,$$

where $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_{K^\star})^T$ is the coefficient for the corresponding $K^\star$ scores of the supervised FPCs, $\hat{\xi}_i = (\hat{\xi}_{i,1}, \hat{\xi}_{i,2}, \ldots, \hat{\xi}_{i,K^\star})^T$. The proportional hazards assumption was deemed reasonable upon inspecting the Schoenfeld residuals plot for each of the baseline covariates. We have plotted the estimated cumulative baseline hazard function for the fitted model using the Breslow estimator, which can be found in Figure S5 in the Supplementary Material. To avoid over-fitting, we have adopted the 10-fold cross-validation in this analysis. For each fold of the cross-validation, the value for θ was chosen from a self-defined grid from 0 to 1 with increment of 0.05 with a nested five-fold cross-validation. The optimal value of θ is chosen to be around 0.1 in most of the cross-validation folds. To choose the number of basis functions $K^\star$, we plotted the objective function $Q(\phi)$ against $K^\star$ and located the "elbow point" where the function flattens out. As shown in Figure 3, the elbow point typically occurs at $K^\star = 2$ in the training folds.

Figure 4 displays the first two FPCs estimated from one of the training samples. The two supervised FPCs retain similar trends as the FPCs estimated from the conventional FPCA, although their values are somewhat different at certain time intervals. For discriminatory measures, the estimated AUC for conventional/supervised FPCA at $t^\star = 6, 12, 18$ with $\triangle t = 6$ are 0.752/0.757, 0.800/0.807, and 0.720/0.734, respectively. The estimated BS for conventional/supervised FPCA are 0.134/0.136, 0.150/0.149, and 0.167/0.169, respectively. The first supervised FPC in Figure 4A is positive with lower values within the time interval [17.0, 42.1] and larger values within the time interval [0.0, 17.0], in comparison with the unsupervised FPC. Because the kth FPC score, $\hat{\xi}_{i,k} = \int_0^\tau \hat{\phi}_k(t)(Z_i(t) - \mu(t))dt$, can be interpreted as the weighted average of the centered ADAS-13, $Z_i(t) - \mu(t)$, in $[0, \tau]$, we can consider the kth FPC $\hat{\phi}_k(t)$ as the time-varying weight on the population level, $k = 1, 2$. Thus, the first supervised FPC suggests a smaller contribution in the time interval [17.0, 42.1] to the risk of AD in comparison to [0.0, 17.0] on average in the population. In

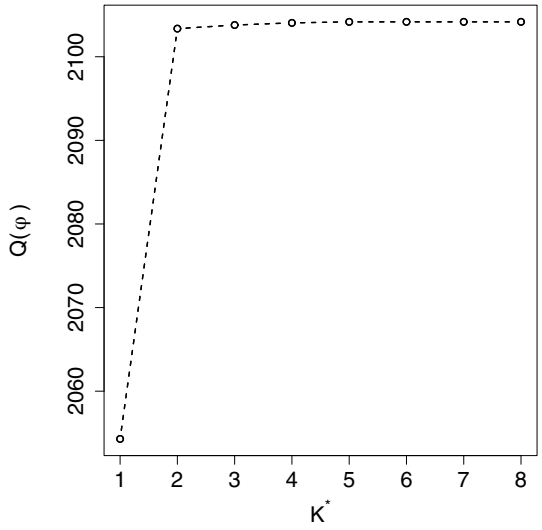

FIGURE 3 The objective function $Q(\phi)$ vs the number of basis functions $K^\star$ in the analysis of ADNI data. The elbow point appear to occur at $K^\star = 2$

general, the positivity of the first supervised FPC seems to be interpretable as an AD-related effect. This interpretation is supported by the individual-specific scores shown in Figure 4C, which are mainly positive for individuals that are diagnosed with AD by their last visit, while scores remained nearly negative for those who remained as MCI. On the other hand, the second FPC in Figure 4B can be interpreted as the variability of population change across time, since the FPCs crossed the zero line. The change in the mean ADAS-13 over time is relative to the baseline ADAS-13 values. Thus, individuals with large baseline ADAS-13 is less likely to have a significant change, compared to individuals who retain relatively smaller baseline values. This interpretation is again, supported by the boxplot of the second individual-specific scores in Figure 4D, which are mainly positive for individuals that are have below average baseline ADAS-13 measurements, while scores are nearly negative for those with above average baseline measurements.

Figure 5 illustrates sample longitudinal path of individuals who retained large and small scores on the first two supervised FPCs. From the left panel, we can see that individual 102 has increasing ADAS-13 during the whole at-risk time interval, while individual 1246 has a relatively flat trajectory. Corresponding to our interpretation on the first FPC score,

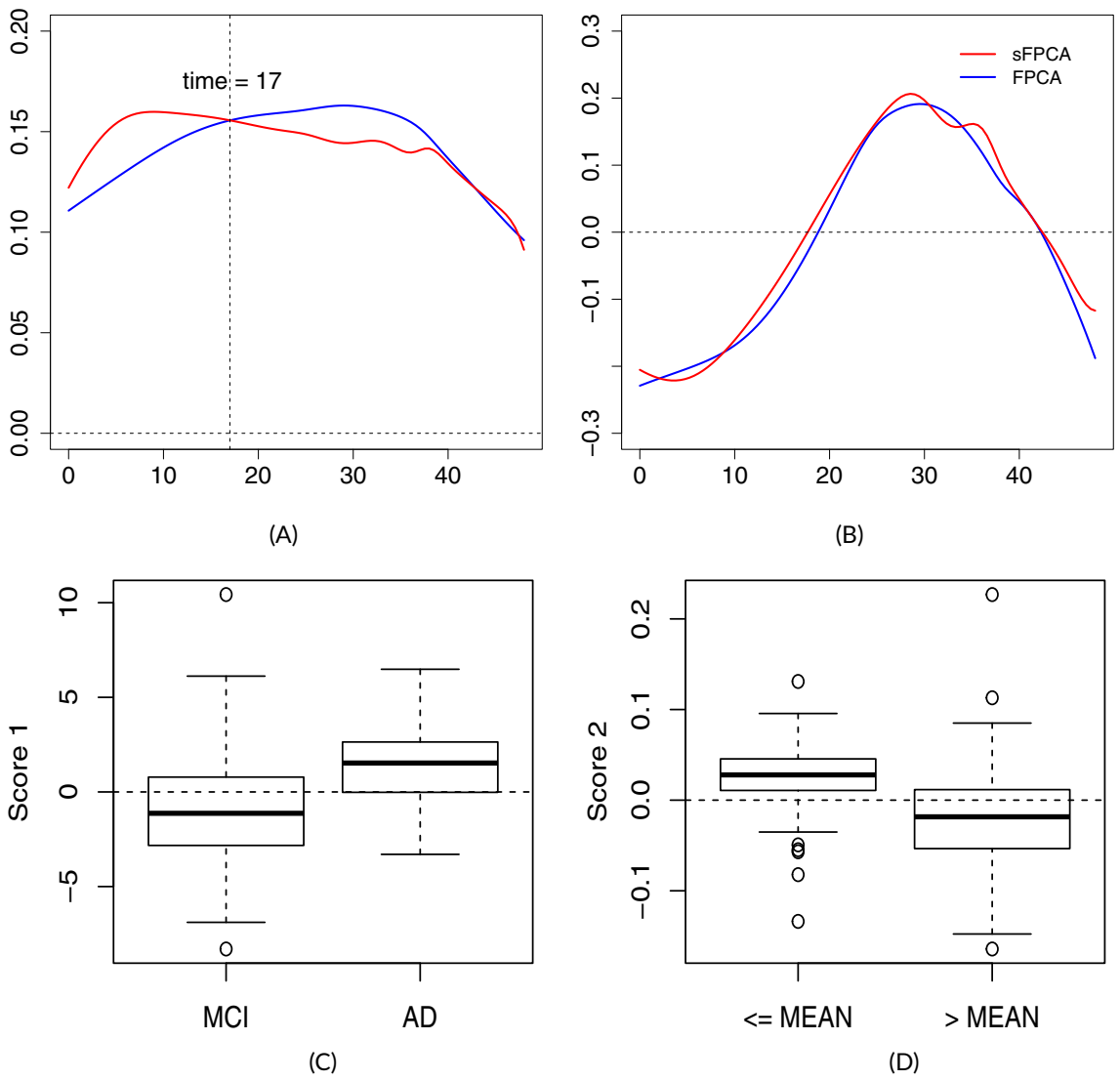

FIGURE 4 The first two functional principal components (FPCs) estimated with conventional FPCA and supervised FPCA (first row); The bottom left panel shows the boxplot of the score on the first FPC on the MCI and AD individuals. The bottom right panel is the boxplot of the score on the second FPC for individuals with the baseline ADAS-13 smaller or larger than the population mean. These results are taken under one of the 10-fold cross-validations. (A) First functional principal component (FPC); (B) second functional principal component (FPC); (C) event status; (D) baseline ADAS-13

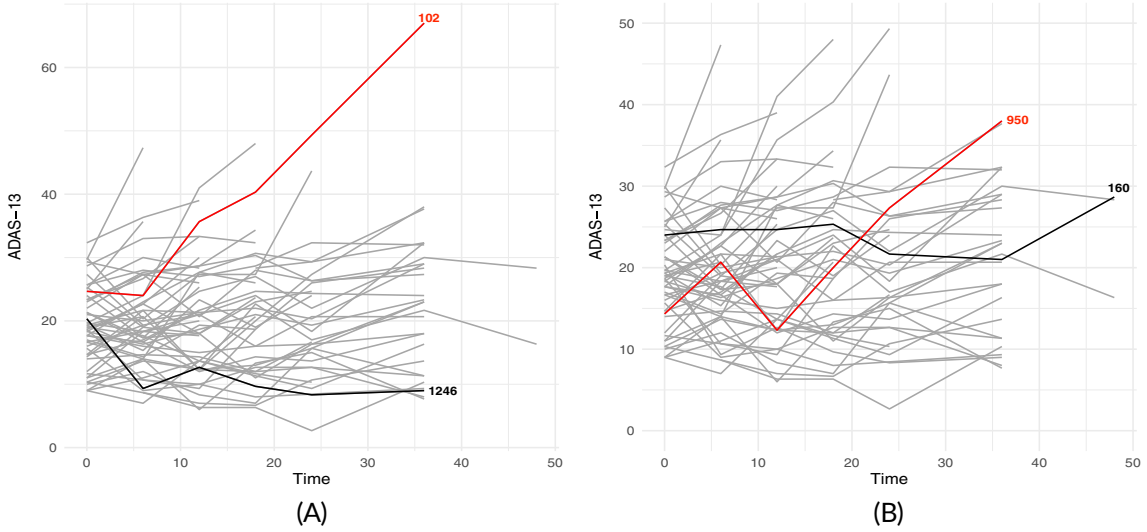

FIGURE 5 An example of individuals who are likely to retain relative large scores (red) and small scores (black) on the first and second functional principal components, shown in the left and right panel, respectively. Individual 102 (score 1 = 9.10, score 2 = 0.17); Individual 1246 (score 1 = −3.76, score 2 = −0.01); Individual 950 (score 1 = 1.45, score 2 = 0.11); Individual 160 (score 1 = 1.62, score 2 = −0.07)

we would expect that individual 102 has a large first score, while individual 1246 will have a small first score, which is indeed the case. Note that individual 1246 has a negative first score, because the centered ADAS-13 values for this subject are negative. The right panel in Figure 5, on the other hand, shows the sample path for individuals 950 and 160, respectively. We can see that individual 950 has a sharp increase after month 13, while individual 160 has a slight increase in ADAS-13 after month 17. We can see that individual 950 has a large second score, 0.11, compared to individual 160 of −0.07. Since our interpretation of the second score is variability of the change across time, such observation is concordant with our interpretation.

We further show in Figure S4 in the Supplemental Material, the scatter plot of scores estimated from the first two supervised FPCs under one of the 10-fold cross-validations. By projecting the individualized longitudinal trajectories onto this two-dimensional space, we are able to identify extreme individuals with ease.

5 | DISCUSSION

The relationship between longitudinal predictor and survival outcomes is important to consider when making clinical decisions. There exists an extensive literature on utilizing FPCA to characterize longitudinal predictors, prior to making individualized predictions. Conventional FPCA-based approaches are all conducted independent of the survival outcomes, and thus the ordering of the FPCs do not indicate the degree of association with the outcome measures. With this consideration, we have proposed a novel supervised FPCA method that can directly estimate a set of basis functions to optimize the association between the longitudinal predictor and the time-to-event outcome. We have investigated the finite sample performance through intensive simulation studies. The simulation results suggest that the proposed method performs well under various scenarios. We have illustrated the utility of the proposed method through an application from the Alzheimer's Disease Neuroimaging Initiative. The estimated set of supervised basis functions provides new insights as discussed in detail under Section 4.

Within the proposed supervised FPCA objective function, estimation for the covariance function between the functional scores and observed time hinges on the assumption of independent censoring. In settings where there is suspected that the censoring mechanism is conditional on the covariate path, our independent censoring assumption can be relaxed to the "uninformative censoring" assumption conditional on the covariates observed.[23] Moreover, while we have considered a single longitudinal predictor in this article, extension to multiple predictors is possible. These will remain as part of our future work.

ACKNOWLEDGEMENTS

This project is partially supported by the NCI (R37 CA256810) and NSERC Discovery Grant (RGPIN-2018-06008). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: http://adni.loni.usc.edu/wpcontent/uploads/howtoapply/ADNIAcknowledgementList.pdf

CONFLICT OF INTEREST

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (http://adni.loni.usc.edu).

ENDNOTE

*http://adni.loni.usc.edu/

ORCID

Haolun Shi https://orcid.org/0000-0002-3620-9717
Shu Jiang https://orcid.org/0000-0003-1464-4838
Jiguo Cao https://orcid.org/0000-0001-7417-6330

REFERENCES

1. Ramsay JO, Silverman BW. *Functional Data Analysis*. 2nd ed. New York, NY: Springer; 2005.
2. Ferraty F, Vieu P. *Nonparametric Functional Data Analysis*. New York, NY: Springer; 2006.
3. Ramsay JO, Dalzell C. Some tools for functional data analysis. *J Royal Stat Soc Ser B*. 1991;53:539-572.
4. Rice JA, Silverman BW. Estimating the mean and covariance structure nonparametrically when the data are curves. *J Royal Stat Soc Ser B*. 1991;53:233-243.
5. Yao F, Muller HG, Wang JL. Functional data analysis for sparse longitudinal data. *J Am Stat Assoc*. 2005;100(470):577-590.
6. Chen K, Lei J. Localized functional principal component analysis. *J Am Stat Assoc*. 2015;110:1266-1275.
7. Lin Z, Wang L, Cao J. Interpretable functional principal component analysis. *Biometrics*. 2016;72:846-854.
8. Nie Y, Cao J. Sparse functional principal component analysis in a new regression framework. *Comput Stat Data Anal*. 2020;152:107016.
9. Yan F, Lin X, Huang X. Dynamic prediction of disease progression for leukemia patients by functional principal component analysis of longitudinal expression levels of an oncogene. *Ann Appl Stat*. 2017;11(3):1649-1670.
10. Wang Y, Ibrahim JG, Zhu H. Partial least squares for functional joint models with applications to the Alzheimer's disease neuroimaging initiative study. *Biometrics*. 2020;76(4):1109-1119. doi:10.1111/biom.13219
11. Jiang S, Xie Y, Colditz GA. Functional ensemble survival tree: dynamic prediction of Alzheimer's disease progression accommodating multiple time-varying covariates. *J Royal Stat Soc Ser C (Appl Stat)*. 2021;70(1):66-79.
12. Li G, Shen H, Huang JZ. Supervised sparse and functional principal component analysis. *J Comput Graph Stat*. 2016;25(3):859-878.
13. Nie Y, Wang L, Liu B, Cao J. Supervised functional principal component analysis. *Stat Comput*. 2018;28(3):713-723.
14. Mercer J. Xvi. functions of positive and negative type, and their connection the theory of integral equations. *Philos Trans Royal Soc Lond Ser A*. 1909;209(441-458):415-446.
15. Ramsay JO, Silverman BW. *Applied Functional Data Analysis: Methods and Case Studies*. New York, NY: Springer; 2007.
16. Welchowski T, Zuber V, Schmid M. Correlation-adjusted regression survival scores for high-dimensional variable selection. *Stat Med*. 2019;38(13):2413-2427.
17. Van der Laan M, Robins JM. *Unified Methods for Censored Longitudinal Data and Causality*. New York, NY: Springer Science & Business Media; 2003.
18. Mardia K, Kent J, Bibby J. *Multivariate Statistics*. New York, NY: Academic Press; 1979.
19. Breslow N. Discussion of the paper by DR. Cox. *J Royal Stat Soc Ser B*. 1972;34(3):216-217.
20. Therneau T, Grambsch P. *Modeling Survival Data: Extending the Cox Model*. New York, NY: Springer; 2000.
21. Rizopoulos D. *Joint Models for Longitudinal and Time-to-Event Data: With Applications in R*. New York, NY: CRC Press; 2012.
22. Van Houwelingen H, Putter H. *Dynamic Prediction in Clinical Survival Analysis*. New York, NY: CRC Press; 2011.
23. Schoop R, Graf E, Schumacher M. Quantifying the predictive performance of prognostic models for censored survival data with time-dependent covariates. *Biometrics*. 2008;64(2):603-613.
24. Hung H, Chiang CT. Estimation methods for time-dependent AUC models with survival data. *Can J Stat*. 2010;38(1):8-26.

25. Blanche P, Proust-Lima C, Loubere L, Berr C, Dartigues FF, Jacqmin-Gadda H. Quantifying and comparing dynamic predictive accuracy of joint models for longitudinal marker and time-to-event in presence of censoring and competing risks. *Biometrics*. 2014;71(1):103-113.
26. Weintraub S, Carrillo MC, Farias ST, et al. Measuring cognition and function in the preclinical stage of Alzheimer's disease. *Alzheimer's Dement Transl Res Clin Intervent*. 2018;4:64-75.

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: Shi H, Jiang S, Cao J. Dynamic prediction with time-dependent marker in survival analysis using supervised functional principal component analysis. *Statistics in Medicine*. 2022;1-14. doi: 10.1002/sim.9433