

ORIGINAL ARTICLE

Tensor Landmark Analysis with application to ADNI data

Sung Hee Park¹  | Ruiwen Zhou¹ | Xin Zhang² | Liang Li³  | Lei Liu¹ ¹Institute for Informatics, Data Science and Biostatistics (I²DB), School of Medicine, Washington University in St. Louis, St. Louis, Missouri, USA²Department of Statistics, Florida State University, Tallahassee, Florida, USA³Department of Biostatistics, MD Anderson Cancer Center, Houston, Texas, USA

Correspondence

Corresponding author Lei Liu, I²DB, School of Medicine, Washington University in St. Louis, St. Louis, MO, USA.
Email: lei.liu@wustl.edu

Funding Information

This research was supported by the National Institutes of Health (NIH) grant R21 AG084054; UL1 TR002345

Abstract

Recent advancements in data collection have facilitated the use of multi-dimensional arrays, also known as tensors, in prediction of health outcomes. In this article, we introduce a tensor landmark model for predicting survival outcomes using multiple longitudinal biomarkers as tensor covariates through CAN-DECOMP/PARAFAC (CP) decomposition. An iteratively reweighted least squares estimation is adopted for components of the tensor coefficients in the landmark model. We also present empirical results of AIC and BIC for right-censored data to select the CP rank. Simulations and Alzheimer's Disease Neuroimaging Initiative (ADNI) data analysis demonstrate that our proposed model accurately estimates survival coefficients and predicts survival probabilities. The implementation code can be found online (<https://github.com/sparkqkr/TensorCoxReg>).

KEYWORDS

survival analysis, landmark analysis, tensor decomposition, dynamic prediction, ADNI data

1 | INTRODUCTION

Alzheimer's disease is a progressive neurodegenerative condition marked by psychiatric, cognitive, and structural deterioration that accounts for 60% to 80% of all dementia cases (Alzheimer's Association, 2021). With no known cure available, there is considerable interest in better understanding the characteristics of the disease. Specifically, early diagnosis is crucial for effective disease management and timely use of disease-modifying drugs.

Our motivating example is the Alzheimer's disease neuroimaging initiative (ADNI) study (Weiner et al., 2010), an international research that actively supports the investigation and development of treatments to alleviate or prevent the progression of Alzheimer's disease. One of the overarching goals of the ADNI study is to detect Alzheimer's disease at the earliest possible stage (pre-dementia) and identify ways to track the progression of the disease. There are numerous literatures on the prediction of participants at risk in an early stage of the disease before symptomatic onset (Mirabnahrzham et al., 2023; Nakagawa et al., 2020; Orozco-Sanchez et al., 2019).

For time-to-event risk prediction, the Cox proportional hazards model (Cox, 1972) is a common approach to handling time-independent risk factors; that is, they do not change over time or are measured at a single time point (e.g., at baseline). When the risk factors vary with time, we can either consider these predictors as time-dependent covariates in a Cox model or adopt the landmark model framework (Anderson et al., 1983). Landmark analysis is desirable when there is a clear rationale for focusing on survival from a specific time point onward, which is often the case when the data are collected at fixed time points, while the Cox model with time-dependent covariates is more suitable when the visiting time is irregular. In the ADNI data, since the visiting schedule is fixed in advance, we will adopt the landmark model framework.

At each landmark time (typically aligned with clinical visiting time), the landmark model predicts the residual lifetime of the at-risk population using covariates defined at the landmark time. This model is thus restricted to the subset of individuals who

survived up to the landmark time. It has been extensively studied in longitudinal research to predict the time-to-event outcome using repeated measured risk factors (L. Li et al., 2017; Van Houwelingen & Putter, 2011; Yao et al., 2023; Zheng & Heagerty, 2005). However, these models typically focus only on covariates measured at or just before the landmark time, disregarding earlier measurements. This approach may overlook important dynamics and relationships among covariates at different times. It is thus imperative to include the history of multivariate time-dependent covariates in landmark analysis.

Data in the form of multi-dimensional arrays, also known as tensors, are becoming increasingly common in a wide range of scientific and clinical applications. We refer the readers to Bi et al. (2021); Mai & Zhang (2023); Sun et al. (2021), which provide comprehensive reviews of the applications of tensor methods. In addition, Zhou et al. (2013) proposed a tensor regression model with a univariate response and a tensor predictor in a generalized linear model (GLM) framework under the CANDECOMP/PARAFAC (CP) decomposition structure. X. Li et al. (2018) extended the GLM based tensor regression to a Tucker decomposition structure. L. Li & Zhang (2017) studied the regression problem with a multi-dimensional array response and a vector predictor. Zhang & Li (2017) proposed a partial least squares (PLS) approach with a vector response and a tensor predictor under the Tucker decomposition structure. Lock (2018) introduced a CP factorized regression model with a tensor covariate and a tensor response. Related work has also explored Bayesian approaches to tensor data, with a specific interest on applications to the ADNI study. Miranda et al. (2018) proposed a Bayesian hierarchical model based on CP decomposition to predict disease status in the ADNI study. Spencer et al. (2022) introduced a Bayesian tensor regression model under Tucker decomposition and compared the prediction outcomes. More recently, Lyu et al. (2024) proposed a data augmentation-based Bayesian classification approach under low-rank CP assumption and showed the classification accuracies on MRI data from the ADNI data.

However, studies on right-censored survival data using tensor longitudinal covariates are limited. In this article, we present a tensor landmark model (TLM), assuming that longitudinal predictors are in a tensor form with a low-rank CP decomposition structure, thereby reducing the number of parameters to be estimated. In the ADNI data, for example, we stack the multiple visits (including the baseline) of 21 longitudinal covariates and treat them as matrix observations. We adopt an estimation approach in Nygård et al. (2008), which updates survival coefficients based on current estimates of the baseline hazard increments using an iteratively reweighted least squares (IRLS) estimation. We extend the estimation approach to the tensor regression coefficients, including the coefficients for the baseline covariate.

The rest of the article is organized as follows. Section 2 introduces the tensor landmark model (TLM). Section 3 presents the estimation procedure with the corresponding algorithm, including the Akaike information criterion (AIC) and the Bayesian information criterion (BIC) for CP rank selection. Section 4 compares our proposed model and the traditional landmark analysis model in the simulation results. Section 5 shows an application of the tensor landmark model (TLM) to the Alzheimer's disease neuroimaging initiative (ADNI) data. We conclude the article with a discussion in Section 6.

2 | MODELS

2.1 | Tensor and CP decomposition

We start with a brief review of tensor notations and operations that will be used in this article. A scalar $a \in \mathbb{R}$, vector $\mathbf{a} \in \mathbb{R}^{p_1}$, matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$ are denoted lowercase, bold lowercase, and bold uppercase, respectively. As mathematical norms within a vector space, for matrix $\mathbf{A}$, $\|\mathbf{A}\|_F$ indicates a Frobenius norm which is a square root of the sum of squared singular values of $\mathbf{A}$, while $\|\mathbf{A}\|_S$ is a spectral norm of $\mathbf{A}$ representing the largest singular value of $\mathbf{A}$. $|a|$ indicates an absolute value of the scalar a . $\mathbf{I}(\cdot)$ is an indicator function, that is, if A is a subset of some set X , then $\mathbf{I}_A(x) = 1$ if $x \in A$, and $\mathbf{I}_A(x) = 0$ otherwise. Lastly, an operator for the unit vector indicates $\text{unit}(\mathbf{a}) = \mathbf{a}/\|\mathbf{a}\|$ where $\|\mathbf{a}\|$ is the size of the vector $\mathbf{a}$.

A D -mode tensor $\mathcal{A} \in \mathbb{R}^{p_1 \times \dots \times p_D}$, denoted by bold calligraphic letters, is a multi-dimensional array that generalizes a matrix (a 2-mode tensor) to higher dimensions. For vectors $\mathbf{a}_1, \dots, \mathbf{a}_D$ of length $I_1, \dots, I_D$, respectively, define the outer product $\mathbf{a}_1 \circ \dots \circ \mathbf{a}_D$ as the D -mode array of dimensions $I_1 \times \dots \times I_D$, with entries $\mathcal{A}_{i_1, \dots, i_D} = \prod_{d=1}^D \mathbf{a}_{d, i_d}$. The $\text{vec}(\cdot)$ operator stacks the entries of the D -dimensional tensor $\mathcal{A} \in \mathbb{R}^{p_1 \times \dots \times p_D}$ into a column vector, so that an entry $\mathcal{A}_{i_1, \dots, i_D}$ maps to the j -th entry of $\text{vec}(\mathcal{A})$ where $j = 1 + \sum_{d=1}^D (i_d - 1) \prod_{d'=1}^{d-1} p_{d'}$. A vectorized inner product $\langle \cdot, \cdot \rangle$ is defined as $\langle \mathcal{A}, \mathcal{B} \rangle = \langle \text{vec}(\mathcal{A}), \text{vec}(\mathcal{B}) \rangle = \sum_{j_1 j_2} a_{j_1 j_2} b_{j_1 j_2}$.

CANDECOMP/PARAFAC (CP) decomposition is a tensor factorization method that generalizes the concept of matrix decomposition to higher-dimensional arrays (tensors). It breaks down a tensor into a sum of rank-one component tensors, where a rank-one tensor is defined as the outer product of vectors. The rank of a tensor $\mathcal{A}$ is the minimum number of rank-one tensors needed to represent $\mathcal{A}$. It is important to note that the definition of rank for a tensor differs from that of a matrix.

For D -mode tensor $\mathcal{A} \in \mathbb{R}^{p_1 \times p_2 \times \dots \times p_D}$, the CP decomposition factorizes the tensor as $\mathcal{A} \approx \sum_{r=1}^R \mathbf{a}_1^r \circ \mathbf{a}_2^r \circ \dots \circ \mathbf{a}_D^r$ where $\mathbf{a}_d^r \in \mathbb{R}^{p_d}$, $d = 1, \dots, D$, and R is the CP rank. Also, when the tensor is a matrix, the CP decomposition is reduced to the singular value decomposition for the matrix. An example of CP decomposition of mode 3 tensor is in Figure 1. See also Kolda & Bader (2009) for more details about tensors.

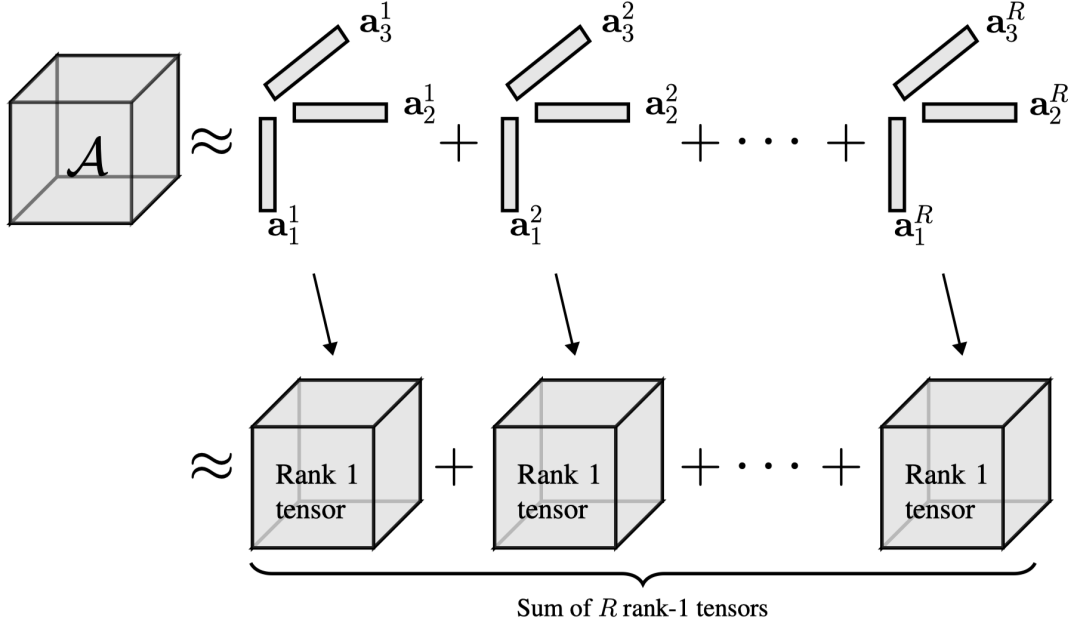


FIGURE 1 Example of CP decomposition of a mode three tensor $\mathcal{A} \in \mathbb{R}^{p_1 \times p_2 \times p_3}$.

2.2 | Landmark model

Landmark analysis was introduced by Anderson et al. (1983). A review was given by Dafni (2011). In landmark analysis, a specific time point s - the landmark time is designated, and the analysis includes only subjects who have survived up to that point. Let $\mathbf{x}_i(s) \in \mathbb{R}^p$ denote a time-dependent observation at time s , where s represents the predefined landmark time. Let $\mathbf{z}_i$ denote a time-independent (baseline) observation. For $t \in (s, s + v)$ where v represent prediction time of clinical interest, a hazard function for each individual i conditional on covariates from time s up to time $s + v$ is

$$\lambda_{\text{LM},i}(t | s) = \lambda_{\text{LM},0}(t | s) \exp \{ \boldsymbol{\eta}^\top \mathbf{z}_i + \boldsymbol{\beta}_{\text{LM}}(s)^\top \mathbf{x}_i(s) \}, \quad i = 1, \dots, n \quad (1)$$

where $\lambda_{\text{LM},0}(t | s)$ is a baseline hazard function at time t for landmark time s , and $\boldsymbol{\beta}_{\text{LM}}(s)$ is a coefficient vector depend on s . The landmark model predicts the risk at s and thus traces the path of the hazard rate from s onward.

2.3 | Tensor landmark model

Assume that we have right-censored survival data with n individuals, denoted by $\{t_i, \delta_i, \mathcal{X}_i, \mathbf{z}_i\}_{i=1, \dots, n}$, where t_i is the observed or censored failure time, δ_i is an indicator of whether the failure time is observed ($\delta_i = 1$) or censored ($\delta_i = 0$), and $\mathbf{z}_i \in \mathbb{R}^q$ is a vector observation of the baseline covariates. $\mathcal{X}_i \in \mathbb{R}^{p_1 \times \dots \times p_D}$ is the observation of a D -mode tensor which stacks $(D - 1)$ -mode tensors by the landmark time s . Thus, the length of the first mode aligns with the number of landmark times, $p_1 = s$. We display the matrix example of the stacked covariates in Figure 2. The tensor landmark model (TLM) is defined as follows:

$$\lambda_{\text{TLM},i}(t | s) = \lambda_{\text{TLM},0}(t | s) \exp \{ \boldsymbol{\eta}^\top \mathbf{z}_i + \langle \mathcal{B}(s), \mathcal{X}_i(s) \rangle \}, \quad i = 1, \dots, n, \quad (2)$$

where $\mathcal{B}(s)$ is a tensor of regression coefficients on the CP decomposition structure as $\mathcal{B}(s) = \sum_{r=1}^R \sigma^r(s) \cdot \langle \mathbf{b}_1^r(s) \circ \dots \circ \mathbf{b}_D^r(s) \rangle \in \mathbb{R}^{p_1 \times \dots \times p_D}$, where $\mathbf{b}_d^r(s) \in \mathbb{R}^{p_d}$ for $d = 1, \dots, D$, and R is the CP rank. We add the scalar parameter $\sigma^r(s) \in \mathbb{R}$ to avoid a scalability issue on each vector $\mathbf{b}_d^r(s)$. Thus, we impose a length constraint on $\mathbf{b}_d^r(s)$. More details are provided in Section 3.1. With appropriate R , we can reduce the number of parameters to $R \times (p_1 + \dots + p_D) + q$ instead of $(p_1 \times \dots \times p_D) + q$. Moreover, model (2) preserves the structural property in $\mathcal{X}_i(s)$ and $\langle \mathcal{B}(s), \mathcal{X}_i(s) \rangle$ summarizes the history of $\mathcal{X}_i(s)$ over landmark times by maintaining the temporal and structural information. We then need to estimate $\mathcal{B}(s)$ under the tensor decomposition structure and $\boldsymbol{\eta}$ without the structural assumption.

To simplify our description, we set the tensor covariates in (2) as a mode 2 tensor, i.e., a matrix, as in the ADNI data example. For the matrix observation $\mathbf{X}_i(s) \in \mathbb{R}^{p_1 \times p_2}$ and the multivariate baseline observation $\mathbf{z}_i$, the hazard function is as follows:

$$\lambda_{\text{TLM},i}(t | s) = \lambda_{\text{TLM},0}(t | s) \exp \{ \boldsymbol{\eta}^\top \mathbf{z}_i + \langle \mathbf{B}(s), \mathbf{X}_i(s) \rangle \}, \quad i = 1, \dots, n, \quad (3)$$

where $\mathbf{B}(s) = \sum_{r=1}^R \sigma^r(s) \cdot \langle \mathbf{b}_1^r(s) \circ \mathbf{b}_2^r(s) \rangle \in \mathbb{R}^{p_1 \times p_2}$ is a matrix of coefficients on the CP decomposition structure where $\mathbf{b}_1^r(s) \in \mathbb{R}^{p_1}$, $\mathbf{b}_2^r(s) \in \mathbb{R}^{p_2}$, $\sigma^r(s) \in \mathbb{R}$, and the CP rank R . For simplicity of notation we omit the subscript "TLM" in the hazard functions hereafter, and use $\lambda(t | s)$ and $\lambda_0(t | s)$ instead.

The full likelihood for tensor landmark model (3) is written as follows:

$$L(\mathbf{B}(s), \boldsymbol{\eta}) = \prod_{i=1}^n [\lambda_0(t_i | s) \exp(\boldsymbol{\eta}^\top \mathbf{z}_i + \langle \mathbf{B}(s), \mathbf{X}_i(s) \rangle)]^{\delta_i} \exp\{-\Lambda_0(t_i | s) \exp(\boldsymbol{\eta}^\top \mathbf{z}_i + \langle \mathbf{B}(s), \mathbf{X}_i(s) \rangle)\},$$

where $\Lambda_0(t_i | s)$ is the cumulative baseline hazard - a non-negative, non-decreasing function. The full likelihood is maximized when all probability masses are concentrated at the distinct observed failure times (Johansen, 1983). With distinct failure times $t_1 < t_2 < \dots < t_W$, the extended full likelihood takes the form of

$$L(\mathbf{B}(s), \boldsymbol{\eta}) = \prod_{i=1}^n [\{\Delta \Lambda_0(t_i | s) \exp(\boldsymbol{\eta}^\top \mathbf{z}_i + \langle \mathbf{B}(s), \mathbf{X}_i(s) \rangle)\}^{\delta_i} \exp\{-\Lambda_0(t_i | s) \exp(\boldsymbol{\eta}^\top \mathbf{z}_i + \langle \mathbf{B}(s), \mathbf{X}_i(s) \rangle)\}], \quad (4)$$

where $\Lambda_0(t_i | s) = \sum_{j: t_j \leq t_i} \Delta \Lambda_0(t_j | s)$, and $\Delta \Lambda_0(t_j | s)$ is the increment of the cumulative baseline hazard at time t_j .

Given values of $\mathbf{B}(s)$ and $\boldsymbol{\eta}$, the baseline hazard increments that maximize likelihood (4) are determined by

$$\Delta \hat{\Lambda}_0(t_j | s) = \frac{d_j}{\sum_{i: t_i \geq t_j} \exp(\boldsymbol{\eta}^\top \mathbf{z}_i + \langle \mathbf{B}(s), \mathbf{X}_i(s) \rangle)}, \quad (5)$$

where $d_j = \sum_{i: t_i = t_j} \delta_i$ is the number of individuals who fail at time t_j . For $t_1 < t_2 < \dots < t_W$, the profile likelihood for the coefficients is

$$PL(\mathbf{B}(s), \boldsymbol{\eta}) = \prod_{j=1}^W \prod_{i: t_i = t_j} \left(\frac{\exp(\boldsymbol{\eta}^\top \mathbf{z}_i + \langle \mathbf{B}(s), \mathbf{X}_i(s) \rangle)}{\sum_{l: t_l \geq t_j} \exp(\boldsymbol{\eta}^\top \mathbf{z}_l + \langle \mathbf{B}(s), \mathbf{X}_l(s) \rangle)} \right)^{\delta_i}. \quad (6)$$

Estimation in the Cox regression model is usually conducted in two steps: first maximize coefficients in profile likelihood (6) using an iterative algorithm, e.g., Newton-Raphson, and then insert the maximum likelihood estimates to the increments of baseline hazard function (5). We have a set of parameters $\boldsymbol{\Theta}(s) = \{\sigma^r(s), \mathbf{b}_1^r(s), \mathbf{b}_2^r(s), \boldsymbol{\eta}, r = 1, \dots, R\}$ where the first three parameters are involved in the tensor coefficient matrix $\mathbf{B}(s)$ for each rank r ; while $\boldsymbol{\eta}$ is not related to r , i.e., it does not depend on the choice of CP rank R . We then adopt a block relaxation algorithm to update the parameter set $\boldsymbol{\Theta}(s)$. Specifically, we update the parameters through an iteratively reweighted least squares (IRLS) estimation for each rank. Nygård et al. (2008) derived the coefficient estimator on the likelihood from the Newton-Raphson algorithm when the baseline hazard increments $\Delta \Lambda_0(t_j)$ are treated as fixed quantities. The details of the estimation procedures are introduced in Algorithm 1.

Taking the ADNI data as an example, in Figure 2, we construct a matrix covariate $\mathbf{X}_i(s) \in \mathbb{R}^{s \times p}$ that stacks the p covariates $\mathbf{x}_i(s)$ for the first s visits. The observation of the stacked matrix variate $\mathbf{X}_i(s)$ has a trajectory of $\mathbf{x}_i(s)$ to predict the risk after s for each subject i . This means that the visiting time s is the same as p_1 in (3), i.e., the number of rows varies by the number of visits for the landmark modeling. Thus, the TLM estimates a different matrix coefficient $\mathbf{B}(s) \in \mathbb{R}^{s \times p}$ for each landmark time s . For ADNI data analysis, we fit the TLM with the first four ($s = 4$), five ($s = 5$), and six ($s = 6$) visits separately.

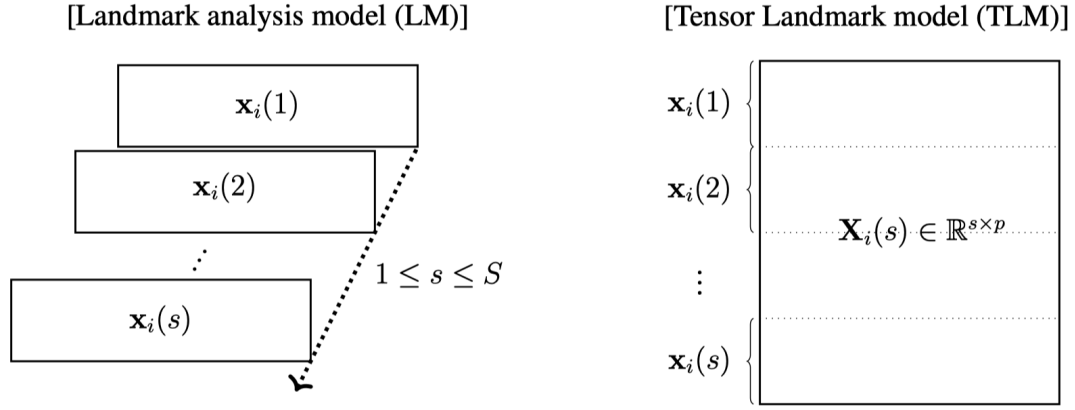


FIGURE 2 Illustration of the covariates in the landmark analysis model (LM) and the tensor landmark model (TLM). In the real data example - the ADNI data, we use 21 longitudinal covariates ($p = 21$) with the first s visits, $s \in \{4, 5, 6\}$.

3 | ESTIMATION

3.1 | Identifiability

For model (3), identifiability for $\mathbf{B}(s)$ implies $\langle \mathbf{X}_i(s), \mathbf{B}(s) \rangle \neq \langle \mathbf{X}_i(s), \mathbf{B}^*(s) \rangle$ if $\mathbf{B}(s) \neq \mathbf{B}^*(s)$. The identifiability is also related to parameters for matrix coefficients $\{\mathbf{b}_1^r(s), \mathbf{b}_2^r(s), \sigma_r(s), r = 1, \dots, R\}$. The conditions required for their identifiability align with those needed for the identifiability of CP decomposition, which has been widely discussed in the literature (Chang et al., 2021; Lock, 2018; Sidiropoulos & Bro, 2000; Zhou et al., 2013). To manage arbitrary scaling and arrangement of components, we impose length constraints on $\mathbf{b}_1^r(s)$ and $\mathbf{b}_2^r(s)$, and the scale will be absorbed in $\sigma_r(s)$. Without loss of generality, we rescale each vector of length 1 to $\tilde{\mathbf{b}}_1^r(s) = \text{unit}(\mathbf{b}_1^r(s)) = \mathbf{b}_1^r(s) / \|\mathbf{b}_1^r(s)\|$ and $\tilde{\mathbf{b}}_2^r(s) = \text{unit}(\mathbf{b}_2^r(s)) = \mathbf{b}_2^r(s) / \|\mathbf{b}_2^r(s)\|$, where $\tilde{\mathbf{b}}_1^r(s)$ and $\tilde{\mathbf{b}}_2^r(s)$ are obtained as the IRLS estimators in Section 3.2 and the unit vectors $\tilde{\mathbf{b}}_1^r(s)$, $\tilde{\mathbf{b}}_2^r(s)$ preserve the directional information.

3.2 | Algorithm

We apply the block relaxation algorithm for the matrix coefficient $\mathbf{B}(s)$ and $\boldsymbol{\eta}$ using the IRLS method in Algorithm 1. To simplify the notation, we omit the landmark time s in $\mathbf{B}(s)$ and $\mathbf{X}(s)$, though a different $\mathbf{B}$ needs to be estimated for each landmark time s . Moreover, we can easily implement this estimation procedure to the Cox proportional hazards model with matrix covariates. The proposed algorithm simultaneously updates the cumulative baseline hazard and the regression coefficient estimates. The inputs are $\{t_i, \delta_i, \mathbf{X}_i, \mathbf{z}_i\}_{i=1, \dots, n}$ and we estimate the matrix coefficients $\mathbf{B} = \sum_{r=1}^R \beta^r = \sum_{r=1}^R \sigma^r \langle \mathbf{b}_1^r \circ \mathbf{b}_2^r \rangle$ under the CP decomposition structure and the baseline coefficient $\boldsymbol{\eta}$. Denote the four components of the parameter set by $\boldsymbol{\Theta} = \{\sigma^r, \mathbf{b}_1^r, \mathbf{b}_2^r, \boldsymbol{\eta}, r = 1, \dots, R\}$. We set the initial values as non-informative values, $\hat{\mathbf{b}}_1^{r(0)} = (0.00001, \dots, 0.00001)^\top \in \mathbb{R}^{p_1}$, $\hat{\mathbf{b}}_2^{r(0)} = (0.00001, \dots, 0.00001)^\top \in \mathbb{R}^{p_2}$, $\hat{\sigma}^{r(0)} = 1$, $\hat{\boldsymbol{\eta}}^{(0)} = \mathbf{0} \in \mathbb{R}^q$, to avoid bias from the starting point. Then we iterate the updating steps until the convergence of the logarithm of full likelihood (4) with the tolerance level ϵ of 10^{-8} .

Algorithm 1 introduces detailed updating steps for the components in $\boldsymbol{\Theta}$. Since the updating procedures for each component in $\boldsymbol{\Theta}$ are similar, we introduce the estimation steps for $\mathbf{b}_1^r$. As a first step, we calculate the pseudo-observation based on the current baseline estimate in the k -th update. The pseudo-observation represents the predicted risk score with current coefficient updates for each individual. In the second step, we formulate the mode-1 observation $\mathbf{x}_{1i}$, which multiplies the other components of the matrix coefficients. We compute the $(k+1)$ -th IRLS update $\tilde{\mathbf{b}}_1^{r(k+1)}$ by using the weight matrix $\hat{\mathbf{V}}_1^{(k)}$ in the third step. Finally, we rescale the $\tilde{\mathbf{b}}_1^{r(k+1)}$ for the unit vector and update $\hat{\mathbf{b}}_1^{r(k+1)}$. The estimation procedures are as follows:

Step 1) Calculate pseudo-observation $\hat{f}_{1i}^{(k)} = \hat{\boldsymbol{\eta}}^{(k)\top} \mathbf{z}_i + \sum_{j=1}^{r-1} \langle \mathbf{X}_i, \hat{\boldsymbol{\beta}}^j \rangle + \langle \mathbf{X}_i, \hat{\boldsymbol{\beta}}^{r(k)} \rangle + (\delta_i - \hat{\mu}_{1i}^{(k)}) / \hat{\mu}_{1i}^{(k)}$, where $\hat{\boldsymbol{\beta}}^j = \hat{\sigma}^j \langle \hat{\mathbf{b}}_1^j \circ \hat{\mathbf{b}}_2^j \rangle$, and $\hat{\mu}_{1i}^{(k)} = \hat{\Lambda}_{10}^{(k)}(t_i) \exp \{ \hat{\boldsymbol{\eta}}^{(k)\top} \mathbf{z}_i + \sum_{j=1}^{r-1} \langle \mathbf{X}_i, \hat{\boldsymbol{\beta}}^j \rangle + \langle \mathbf{X}_i, \hat{\boldsymbol{\beta}}^{r(k)} \rangle \}$ when $\hat{\Lambda}_{10}^{(k)}(t_i)$ is a cumulative Breslow baseline hazard (Lin, 2007) at time t_i with current updates $\hat{\boldsymbol{\beta}}^{r(k)}$ and $\hat{\boldsymbol{\eta}}^{(k)}$.

- Step 2) For the least-squares approach, compute the mode-1 observation $\mathbf{x}_{1i} = \hat{\sigma}^{r(k)} \mathbf{X}_i \hat{\mathbf{b}}_2^{r(k)}$. Then, the mode-1 observation matrix is $\mathbf{X}_1 = (\mathbf{x}_{1i}, \dots, \mathbf{x}_{1n})^\top \in \mathbb{R}^{n \times p_1}$.
- Step 3) $\hat{\mathbf{b}}_1^{r(k+1)} = (\mathbf{X}_1^\top \hat{\mathbf{V}}_1^{(k)} \mathbf{X}_1)^{-1} \mathbf{X}_1^\top \hat{\mathbf{V}}_1^{(k)} \hat{\mathbf{f}}_1^{(k)}$ where $\hat{\boldsymbol{\mu}}_1^{(k)} = (\hat{\mu}_{11}^{(k)}, \dots, \hat{\mu}_{1n}^{(k)})^\top$, $\hat{\mathbf{V}}_1^{(k)} = \text{diag}(\hat{\boldsymbol{\mu}}_1^{(k)})$, and the vector of pseudo-observations $\hat{\mathbf{f}}_1^{(k)} = (\hat{f}_{1i}^{(k)}, \dots, \hat{f}_{1n}^{(k)})^\top$.
- Step 4) Update $\hat{\mathbf{b}}_1^{r(k+1)}$ as the unit vector that is calculated as $\hat{\mathbf{b}}_1^{r(k+1)} = \text{unit}(\hat{\mathbf{b}}_1^{r(k+1)})$.

In Algorithm 1, $\mathbf{b}_2^r$ has similar estimation steps to $\mathbf{b}_1^r$. The scalar parameter σ^r and the clinical coefficient $\boldsymbol{\eta}$ do not require the last step (step 4) since the unit vector constraint is for the identification of the factors in the CP decomposition, as described in Section 3.1. Nygård et al. (2008) showed that this least squares estimator can be obtained from the Newton-Raphson algorithm using a score function of full likelihood (4).

3.3 | Rank selection

One major challenge is that there is no straightforward algorithm to determine the rank of a given specific tensor; in fact, the problem is NP-hard (Håstad, 1989). We implement the Akaike information criterion (AIC; Akaike, 1974) and the Bayesian information criterion (BIC; Schwarz, 1978) for rank selection. Liang & Zou (2008) proposed an improved AIC variable selection method for survival analysis. Volinsky & Raftery (2000) suggested a modification to the penalty term in the BIC, defining it based on the number of uncensored events rather than the total number of observations. Here, the AIC and BIC with right-censored data are as follows:

$$\text{AIC} = -2L(\mathbf{B}, \boldsymbol{\eta}) + 2d_R, \quad \text{BIC} = -2L(\mathbf{B}, \boldsymbol{\eta}) + d_R \log \left(\sum_{i=1}^n \delta_i \right), \quad (7)$$

where $d_R = R(p_1 + p_2) + q$ represents the degree of freedom under rank R and δ_i is event for each individual. We show how both AIC and BIC work on different strength of signals in Table 2.

4 | SIMULATION

In this section, we examine the estimation of coefficients in the tensor landmark model (TLM). We first set $p_1 = 8$, $p_2 = 9$, $q = 2$ and generate matrix variate observations $\mathbf{X}_i \sim \mathcal{MN}(\mathbf{0}, \boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2)$, $i = 1, \dots, n$ where $\mathcal{MN}(\cdot)$ represents a matrix normal distribution with matrix mean $\mathbf{0}$ and covariances $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_2$, and the covariances have autoregressive structure with parameter 0.6 in each mode. A definition of matrix normal distribution is introduced in the Supplementary Material A. We set $\mathbf{z} = (z_1, z_2)$ where z_1 follows a binomial distribution with a probability of 0.5 and z_2 follows a standard normal distribution.

We consider two different signal shapes on $\mathbf{B} \in \mathbb{R}^{8 \times 9}$ for the tensor landmark model (TLM): some variables are available at all visits as the vertical shape in panel (a) in Figure 3, and some variables are significant for a few visits as the square shape in panel (e) in Figure 3. The positions of the non-zero values are:

- Vertical shape (I): values at 1 in the fourth and fifth columns,
- Square shape (■): values at 1 in the upper-left block with size 4×4 .

We can check the non-zero coefficients of $\mathbf{B}$ as a colored part in the panels (a) and (e) in Figure 3. The coefficients for $\mathbf{z}$ are set as $\boldsymbol{\eta} = (0.8, 1.2)$. For survival data generation, we assume that the baseline hazard follows a Weibull distribution with scale and shape parameters $(\alpha, \nu) = (2, 3)$. The survival time is then generated using $t_i = [-\alpha \log(U_i) / \exp(\boldsymbol{\eta}^\top \mathbf{z}_i + \langle \mathbf{B}, \mathbf{X}_i \rangle)]^\nu$, where U_i is drawn from the uniform distribution range of $[0, 1]$. The censoring time C_i is generated from a uniform distribution range of $[0, 1]$, resulting in a censoring rate of approximately 50%.

To assess the performance of estimation for $\mathbf{B}$, we use a relative Frobenius norm loss and a relative spectral norm loss, denoted by loss_F and loss_S , respectively. We divide the norm losses by the corresponding norm of the true coefficient $\mathbf{B}$ in the range of 0 to 1. When the norm loss is 1, we have $\hat{\mathbf{B}} = \mathbf{0}$, and the estimation is meaningless. The effective estimation has a relative norm

Algorithm 1 Block relaxation algorithm for tensor landmark model (TLM)

Input: $\{t_i, \delta_i, \mathbf{X}_i, \mathbf{z}_i\}_{i=1}^n, R$
Output: $\hat{\mathbf{B}} = \sum_{r=1}^R \hat{\beta}^r = \sum_{r=1}^R \hat{\sigma}^r \cdot \langle \hat{\mathbf{b}}_1^r \circ \hat{\mathbf{b}}_2^r \rangle, \hat{\eta}$

- 1: **Initialize** $\hat{\eta}^{(0)}$
- 2: **for** r in $1, \dots, R$ **do**
- 3: **Initialize** $\hat{\mathbf{b}}_1^{(0)}, \hat{\mathbf{b}}_2^{(0)}, \hat{\sigma}^{(0)}$
- 4: **repeat** $k=0, 1, \dots$
- 5: Calculate $\hat{f}_{1i}^{(k)} = \hat{\eta}^{(k)\top} \mathbf{z}_i + \sum_{j=1}^{r-1} \langle \mathbf{X}_i, \hat{\beta}^j \rangle + \langle \mathbf{X}_i, \hat{\beta}^{r(k)} \rangle + \frac{(\delta_i - \hat{\mu}_{1i}^{(k)})}{\hat{\mu}_{1i}^{(k)}} \quad \triangleright$ **Update for** $\hat{\mathbf{b}}_1^r$
 where $\hat{\beta}^j = \hat{\sigma}^j \cdot \langle \hat{\mathbf{b}}_1^j \circ \hat{\mathbf{b}}_2^j \rangle$, $\hat{\mu}_{1i}^{(k)} = \hat{\Lambda}_{10}^{(k)}(t_i) \exp \{ \hat{\eta}^{(k)\top} \mathbf{z}_i + \sum_{j=1}^{r-1} \langle \mathbf{X}_i, \hat{\beta}^j \rangle + \langle \mathbf{X}_i, \hat{\beta}^{r(k)} \rangle \}$
 when $\hat{\Lambda}_{10}^{(k)}(t_i)$ is a cumulative Breslow baseline hazard at time t_i .
- 6: Calculate the mode-1 observation $\mathbf{x}_{1i} = \hat{\sigma}^{r(k)} \cdot \mathbf{X}_i \hat{\mathbf{b}}_2^{r(k)}$, $i = 1, \dots, n$
- 7: Calculate $\tilde{\mathbf{b}}_1^{r(k+1)} = (\mathbf{X}_1^\top \hat{\mathbf{V}}_1^{(k)} \mathbf{X}_1)^{-1} \mathbf{X}_1 \hat{\mathbf{V}}_1^{(k)} \hat{\mathbf{f}}_1^{(k)}$
 where $\hat{\mu}_1^{(k)} = (\hat{\mu}_{11}^{(k)}, \dots, \hat{\mu}_{1n}^{(k)})^\top$, $\hat{\mathbf{V}}_1^{(k)} = \text{diag}(\hat{\mu}_1^{(k)})$, $\hat{\mathbf{f}}_1^{(k)} = (\hat{f}_{11}^{r(k)}, \dots, \hat{f}_{1n}^{r(k)})^\top$.
- 8: Update $\hat{\mathbf{b}}_1^{r(k+1)} \leftarrow \text{unit}(\tilde{\mathbf{b}}_1^{r(k+1)})$
- 9: Calculate $\hat{f}_{2i}^{(k)} = \hat{\eta}^{(k)\top} \mathbf{z}_i + \sum_{j=1}^{r-1} \langle \mathbf{X}_i, \hat{\beta}^j \rangle + \langle \mathbf{X}_i, \hat{\beta}_2^{r(k)} \rangle + \frac{(\delta_i - \hat{\mu}_{2i}^{(k)})}{\hat{\mu}_{2i}^{(k)}} \quad \triangleright$ **Update for** $\hat{\mathbf{b}}_2^r$
 where $\hat{\beta}_2^{r(k)} = \hat{\sigma}^{r(k)} \cdot \hat{\mathbf{b}}_1^{r(k+1)} \circ \hat{\mathbf{b}}_2^{r(k)}$, $\hat{\mu}_{2i}^{(k)} = \hat{\Lambda}_{20}^{(k)}(t_i) \exp \{ \hat{\eta}^{(k)\top} \mathbf{z}_i + \sum_{j=1}^{r-1} \langle \mathbf{X}_i, \hat{\beta}^j \rangle + \langle \mathbf{X}_i, \hat{\beta}_2^{r(k)} \rangle \}$
 when $\hat{\Lambda}_{20}^{(k)}(t_i)$ is a cumulative Breslow baseline hazard at time t_i .
- 10: Calculate the mode-2 observation $\mathbf{x}_{2i} = \hat{\sigma}^{r(k)} \cdot \mathbf{X}_i^\top \hat{\mathbf{b}}_1^{r(k+1)}$, $i = 1, \dots, n$
- 11: Calculate $\tilde{\mathbf{b}}_2^{r(k+1)} = (\mathbf{X}_2^\top \hat{\mathbf{V}}_2^{(k)} \mathbf{X}_2)^{-1} \mathbf{X}_2 \hat{\mathbf{V}}_2^{(k)} \hat{\mathbf{f}}_2^{(k)}$
 where $\hat{\mu}_2^{(k)} = (\hat{\mu}_{21}^{(k)}, \dots, \hat{\mu}_{2n}^{(k)})^\top$, $\hat{\mathbf{V}}_2^{(k)} = \text{diag}(\hat{\mu}_2^{(k)})$, $\hat{\mathbf{f}}_2^{(k)} = (\hat{f}_{21}^{r(k)}, \dots, \hat{f}_{2n}^{r(k)})^\top$.
- 12: Update $\hat{\mathbf{b}}_2^{r(k+1)} \leftarrow \text{unit}(\tilde{\mathbf{b}}_2^{r(k+1)})$
- 13: Calculate $\hat{f}_{3i}^{(k)} = \hat{\eta}^{(k)\top} \mathbf{z}_i + \sum_{j=1}^{r-1} \langle \mathbf{X}_i, \hat{\beta}^j \rangle + \langle \mathbf{X}_i, \hat{\beta}_3^{r(k)} \rangle + \frac{(\delta_i - \hat{\mu}_{3i}^{(k)})}{\hat{\mu}_{3i}^{(k)}} \quad \triangleright$ **Update for** $\hat{\sigma}^r$
 where $\hat{\beta}_3^{r(k)} = \hat{\sigma}^{r(k)} \cdot \hat{\mathbf{b}}_1^{r(k+1)} \circ \hat{\mathbf{b}}_2^{r(k+1)}$, $\hat{\mu}_{3i}^{(k)} = \hat{\Lambda}_{30}^{(k)}(t_i) \exp \{ \hat{\eta}^{(k)\top} \mathbf{z}_i + \sum_{j=1}^{r-1} \langle \mathbf{X}_i, \hat{\beta}^j \rangle + \langle \mathbf{X}_i, \hat{\beta}_3^{r(k)} \rangle \}$
 when $\hat{\Lambda}_{30}^{(k)}(t_i)$ is a cumulative Breslow baseline hazard at time t_i .
- 14: Calculate the scalar observation $x_{3i} = \hat{\sigma}^{r(k)} \cdot \hat{\mathbf{b}}_1^{r(k+1)} \mathbf{X}_i \hat{\mathbf{b}}_2^{r(k+1)}$, $i = 1, \dots, n$
- 15: Update $\hat{\sigma}^{r(k+1)} \leftarrow (\mathbf{x}_3^\top \hat{\mathbf{V}}_3^{(k)} \mathbf{x}_3)^{-1} \mathbf{x}_3 \hat{\mathbf{V}}_3^{(k)} \hat{\mathbf{f}}_3^{(k)}$
 where $\hat{\mu}_3^{(k)} = (\hat{\mu}_{31}^{(k)}, \dots, \hat{\mu}_{3n}^{(k)})^\top$, $\hat{\mathbf{V}}_3^{(k)} = \text{diag}(\hat{\mu}_3^{(k)})$, $\hat{\mathbf{f}}_3^{(k)} = (\hat{f}_{31}^{r(k)}, \dots, \hat{f}_{3n}^{r(k)})^\top$.
- 16: Calculate $\hat{f}_{4i}^{(k)} = \hat{\eta}^{(k)\top} \mathbf{z}_i + \sum_{j=1}^{r-1} \langle \mathbf{X}_i, \hat{\beta}^j \rangle + \langle \mathbf{X}_i, \hat{\beta}_4^{r(k)} \rangle + \frac{(\delta_i - \hat{\mu}_{4i}^{(k)})}{\hat{\mu}_{4i}^{(k)}} \quad \triangleright$ **Update for** $\hat{\eta}$
 where $\hat{\beta}_4^{r(k)} = \hat{\sigma}^{r(k+1)} \cdot \hat{\mathbf{b}}_1^{r(k+1)} \circ \hat{\mathbf{b}}_2^{r(k+1)}$, $\hat{\mu}_{4i}^{(k)} = \hat{\Lambda}_{40}^{(k)}(t_i) \exp \{ \hat{\eta}^{(k)\top} \mathbf{z}_i + \sum_{j=1}^{r-1} \langle \mathbf{X}_i, \hat{\beta}^j \rangle + \langle \mathbf{X}_i, \hat{\beta}_4^{r(k)} \rangle \}$
 when $\hat{\Lambda}_{40}^{(k)}(t_i)$ is a cumulative Breslow baseline hazard at time t_i .
- 17: Update $\hat{\eta}^{(k+1)} \leftarrow (\mathbf{Z}^\top \hat{\mathbf{V}}_4^{(k)} \mathbf{Z})^{-1} \mathbf{Z} \hat{\mathbf{V}}_4^{(k)} \hat{\mathbf{f}}_4^{(k)}$
 where $\mathbf{Z} = (\mathbf{z}_1^\top, \dots, \mathbf{z}_n^\top)^\top$, $\hat{\mu}_4^{(k)} = (\hat{\mu}_{41}^{(k)}, \dots, \hat{\mu}_{4n}^{(k)})^\top$, $\hat{\mathbf{V}}_4^{(k)} = \text{diag}(\hat{\mu}_4^{(k)})$, $\hat{\mathbf{f}}_4^{(k)} = (\hat{f}_{41}^{r(k)}, \dots, \hat{f}_{4n}^{r(k)})^\top$.
- 18: **until convergence** $|\log L(\hat{\Theta}^{(k+1)}) - \log L(\hat{\Theta}^{(k)})| < \epsilon$
- 19: **end for**

loss of less than 1 and close to 0. Relative norm losses are defined as

$$\text{loss}_F = \frac{\|\mathbf{B} - \hat{\mathbf{B}}\|_F}{\|\mathbf{B}\|_F}, \quad \text{loss}_S = \frac{\|\mathbf{B} - \hat{\mathbf{B}}\|_S}{\|\mathbf{B}\|_S},$$

where $\|\cdot\|_F$ indicates a Frobenius norm and $\|\cdot\|_S$ represents a spectral norm. The performance of the estimates for coefficients $\eta = (\eta_1, \eta_2)$ is evaluated by the mean absolute bias and the mean squared error (MSE). Additionally, we calculate the bias as a ratio of the true value, referred to as 'RBias.' A lower RBias, closer to 0, indicates better estimation accuracy.

We compare the tensor landmark model (TLM) with a standard landmark model that vectorizes the matrix covariate $\mathbf{X}_i$ (LM-Vec) in Table 1. For LM-Vec, we concatenate each row of $\mathbf{X}_i$ in a multivariate form (as in Figure 2), and use the R package *survival* (Therneau et al., 2000). We select the CP rank as 2 which has the lowest value of AIC. Thus, the number of parameters is reduced from $8 \times 9 + 2 = 74$ to $R \times (p_1 + p_2) + q = 2(8 + 9) + 2 = 36$. As in Figure 3, the non-zero coefficients (signals) are recovered at rank 2. For regression coefficients $\boldsymbol{\eta} = (\eta_1, \eta_2)^\top$, the TLM has lower bias and MSE in different sample sizes and is more accurate than the LM-Vec regression model. A clearer difference is evident in the relative norm losses of the matrix coefficient $\mathbf{B}$. When the sample size (n) is 300, the LM-Vec model shows relative norm losses (loss_F , loss_S) close to 1, indicating that it fails to capture the signals in $\mathbf{B}$. In contrast, the relative norm losses of the TLM (≤ 0.2) demonstrate that it accurately captures the structured signals in $\mathbf{B}$. Furthermore, as the sample size increases, all values are consistently decreased. Thus, the results in Table 1 support that the TLM is more efficient and accurate than the LM-Vec model.

Shape	n	Model	Parameter					
			η_1		η_2		$\mathbf{B}$	
			RBias	MSE	RBias	MSE	loss_F	loss_S
I	300	TLM	0.2331 (0.0284)	0.0857	0.2176 (0.0125)	0.0903	0.1960 (0.0048)	0.1703 (0.0047)
		LM-Vec	0.6090 (0.0769)	0.6121	1.3172 (0.0423)	1.1025	0.9172 (0.0438)	1.0344 (0.0295)
	600	TLM	0.0943 (0.0216)	0.0354	0.0989 (0.0081)	0.0234	0.1389 (0.0034)	0.1203 (0.0032)
		LM-Vec	0.2435 (0.0309)	0.0984	0.2044 (0.0123)	0.0817	0.5027 (0.0085)	0.3291 (0.0065)
	900	TLM	0.0261 (0.0148)	0.0142	0.0362 (0.0071)	0.0090	0.1278 (0.0027)	0.1166 (0.0026)
		LM-Vec	0.1674 (0.0170)	0.0362	0.1446 (0.0091)	0.0419	0.3417 (0.0040)	0.2272 (0.0033)
■	300	TLM	0.1994 (0.0253)	0.0657	0.2237 (0.0104)	0.0875	0.1876 (0.0040)	0.1580 (0.0037)
		LM-Vec	1.0019 (0.0884)	1.1367	0.7519 (0.0421)	1.0670	1.5354 (0.0467)	0.9931 (0.0315)
	600	TLM	0.1266 (0.0200)	0.0355	0.1018 (0.0074)	0.0228	0.1284 (0.0030)	0.1093 (0.0026)
		LM-Vec	0.1911 (0.0261)	0.0668	0.2281 (0.0106)	0.0910	0.5016 (0.0077)	0.3264 (0.0057)
	900	TLM	0.0603 (0.0121)	0.0169	0.0669 (0.0078)	0.0124	0.1087 (0.0023)	0.0933 (0.0023)
		LM-Vec	0.1596 (0.0164)	0.0428	0.1353 (0.0096)	0.0355	0.3430 (0.0058)	0.2263 (0.0044)

TABLE 1 Numerical assessments of parameter estimation with 100 replicates. The true value is $(\eta_1, \eta_2) = (0.8, 1.2)$. loss_F and loss_S indicate relative Frobenius and spectral norm losses, respectively. RBias represents a ratio of the bias and the true value. The numbers in parentheses represent the standard errors (SE).

We also consider the strength of the signals (values of non-zero coefficients in $\mathbf{B}$) in three magnitudes: 0.3 (weak), 1 (moderate) and 3 (strong). Table 2 shows the ratio of selected rank by AIC and BIC with 100 replications. We find that the AIC and BIC results in the three different magnitudes of signals are mostly consistent.

5 | REAL DATA ANALYSIS

Alzheimer’s disease (AD) is a complex neurodegenerative disorder marked by progressive loss of memory and cognitive functions. It is the leading cause of dementia, projected to affect more than 78 million people by 2030 (Gauthier et al., 2021). The Alzheimer’s disease neuroimaging initiative (ADNI), launched in 2004, is an ongoing research project that aims to develop clinical, imaging, genetic, and biochemical biomarkers for the early detection and monitoring of Alzheimer’s disease progression (Weiner et al., 2010). As a multi-site longitudinal study, ADNI tracks 2,428 individuals (by April 2023) aged between 55 and 90 at sixty three sites in the US and Canada. During repeated intervals over subsequent years, participants underwent clinical

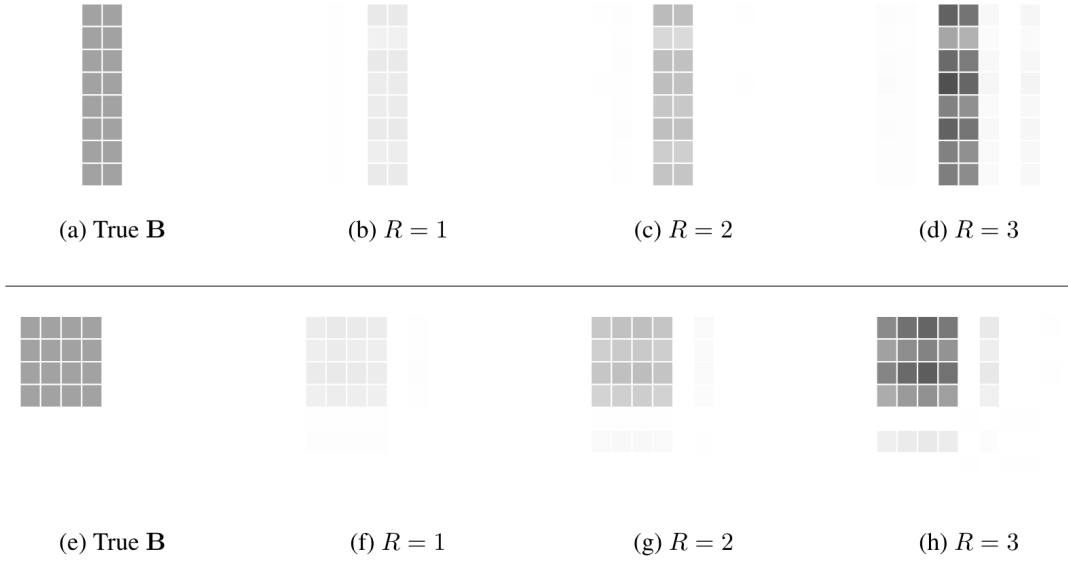


FIGURE 3 True and estimated matrix coefficient by rank in different signal patterns (I, ■) with sample size $n = 600$.

Shape	Magnitude	n	$\hat{R}_{AIC}$			$\hat{R}_{BIC}$		
			1	2	3	1	2	3
I	Weak	300	1.00	0	0	1.00	0	0
		600	1.00	0	0	1.00	0	0
		900	1.00	0	0	1.00	0	0
	Moderate	300	0	1.00	0	0.03	0.97	0
		600	0	1.00	0	0.02	0.98	0
		900	0	1.00	0	0	1.00	0
	Strong	300	0	0	1.00	0	0	1.00
		600	0	0.01	0.99	0	0.01	0.99
		900	0	0	1.00	0	0	1.00
■	Weak	300	1.00	0	0	1.00	0	0
		600	1.00	0	0	1.00	0	0
		900	1.00	0	0	1.00	0	0
	Moderate	300	0	1.00	0	0.01	0.99	0
		600	0	1.00	0	0	1.00	0
		900	0	1.00	0	0	1.00	0
	Strong	300	0	0	1.00	0	0	1.00
		600	0	0	1.00	0	0	1.00
		900	0	0	1.00	0	0	1.00

TABLE 2 Rank selection on the AIC and BIC in different shape and magnitude of signals: weak (non-zero values in $\mathbf{B}$ is 0.3), moderate (non-zero values in $\mathbf{B}$ is 1), strong (non-zero values in $\mathbf{B}$ is 3) with 100 replicates. The numbers are the percentage of each rank selected by AIC and BIC, $\hat{R}_{AIC}$ and $\hat{R}_{BIC}$, respectively. The bolded number represents the most selected rank $\hat{R}$.

evaluation, neuropsychological tests, genetic tests, lumbar puncture, and magnetic resonance and positron emission tomography scans. During each visit, ADNI participants were classified as cognitively normal (CN), mildly cognitively impaired (MCI), or affected by dementia.

In this section, we apply the tensor landmark model (TLM) to data from Alzheimer's disease neuroimaging initiative (ADNI). Here, we predict the time from the landmark time to the diagnosis of dementia for subjects who entered the study as CN or MCI. For analysis, we adopt five baseline covariates (age at baseline, sex, baseline diagnosis, number of apolipoprotein alleles $\epsilon 4$, and years of education) and 21 longitudinal covariates listed in the Supplementary Table C1. These covariates are chosen based on their relevance as dementia predictors and their relatively low incidence of missing values. Missing values are imputed using the most recently recorded value of that variable. Due to significant skewness and differing scales among some variables,

we apply logarithmic transformations to seven variables: entorhinal, fusiform, hippocampus, ICV, whole brain, ventricles, and midtemporal. To train the models, we use 5-fold cross validation (5-fold cv). In each fold, we ensure that the distribution of patient events mirrors that of the original data set by applying stratified sampling to the ADNI data. This approach maintains the same ratio of censored and uncensored patients in each fold.

We compare our proposed tensor landmark model (TLM), the landmark model with the covariate at the landmark time (LM; Anderson et al. 1983), and the landmark model with vectorized longitudinal covariates (LM-Vec). The prediction performance is evaluated using the concordance index (C-index) (Harrell et al., 1982) and the integrated inverse probability of censoring weighting (IPCW) Brier score (IBS). The C-index, an extension of the area under the receiver operating characteristic curve (AUC) for right-censored survival data, ranges from 0 to 1, where 1 indicates perfect prediction and 0.5 indicates a random ordering of risk scores. The Brier score (Brier, 1950) measures the accuracy of probabilistic predictions and has been adapted for right-censored data by Gerds & Schumacher (2006) and Graf et al. (1999) as the IPCW Brier score. To summarize prediction accuracy over multiple time points, we integrate the IPCW Brier scores for 2-year predictions from each landmark time s . The IBS ranges from 0 to 1, with lower values indicating better prediction performance. Further details on the C-index and integrated IPCW Brier score are provided in the Supplementary Material B.

The results are tabulated in Table 3. We set the CP rank to 1 for TLM. Thus, compared to the LM-Vec model, TLM reduces the number of parameters from (89,110,131) to (30,31,32) for each visit $s \in \{4, 5, 6\}$. In Table 3, the TLM achieves the highest C-indexes and the lowest IBS at three different landmark times, indicating that matrix risk factor history from baseline to the s -th visit improves the prediction performance. In contrast, the LM-Vec method shows the lowest C-indexes since the concatenated multivariate risk factors at different time points may have multicollinearity issues.

Year	Number of visits (s)	Status		Sample size	Model	Assessment	
		0	1			C-index	IBS
1.5	4	556	69	625	TLM	0.9180	0.0382
					LM	0.9154	0.0430
					LM-Vec	0.8337	0.0534
2	5	516	27	543	TLM	0.8883	0.0226
					LM	0.8757	0.0263
					LM-Vec	0.7589	0.0250
2.5	6	428	26	454	TLM	0.9238	0.0164
					LM	0.9123	0.0187
					LM-Vec	0.8008	0.0178

TABLE 3 Assessments for the predictions on the year 1.5, 2 and 2.5 in 5-fold cross validation (5-fold cv).

6 | DISCUSSION

In this article, we have proposed the tensor landmark model under a low-rank tensor structure. The numerical results demonstrated that the proposed IRLS solution performs well in both estimation and prediction. We make a few remarks regarding our proposed models. First, our results are also aligned with the fact that the estimation of profile likelihood alone may not be sufficient, as shown in Nygård et al. (2008). Moreover, we can maintain structural information in the tensor covariate using the block relaxation IRLS algorithm. Second, the tensor landmark model can convey the trajectory information of risk factors from baseline to the landmark time, incorporating the longitudinal history prior to the landmark time. Third, the AIC and BIC for rank selection consistently estimate the CP rank when the signals vary in magnitude. In practice, when we have enough samples, the rank one or two is enough for survival prediction.

Our work also suggests several possible extensions. First, image data can be considered as a matrix of covariates in the tensor landmark model. Next, instead of the CP decomposition for low-rank structures, a Tucker decomposition could be naturally imposed on the matrix/tensor covariates. Under the Tucker assumption, we could effectively estimate the matrix/tensor coefficient by imposing useful constraints, e.g., orthogonality assumptions on factor matrices. Moreover, envelope modeling could be applied to the tensor landmark model. Cook & Zhang (2015) introduced a general constructive framework by applying the sufficient dimension reduction technique to the Cox regression model. Relevant information for estimating the survival

coefficients in the envelope modeling gives more efficient estimates of the regression coefficients in the tensor landmark model. Finally, covariates at the latest visiting time might be closely related to the prediction time and may be the most predictive of residual survival time. We could effectively estimate the matrix coefficient by selecting the significant rows (which is similar to the variable selection for all visits).

AUTHOR CONTRIBUTIONS

SP and L Liu envisioned the presented idea. SP developed the models and performed the computations under the supervision of L Liu. RZ furnished the data. XZ, RZ, and L Li verified the analytical methods and provided comments. L Liu provided funding support.

ACKNOWLEDGMENTS

This research is supported by NIH grant R21 AG084054; UL1 TR002345. Data used in the preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (<http://adni.loni.usc.edu>). The investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in the analysis or writing of this article. A complete list of ADNI investigators can be found at: https://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_DSP_Policy.pdf

CONFLICT OF INTEREST

Dr. Lei Liu is consultant to Adial Pharmaceuticals, and Lecturer for Pharmaliaison and the Academy of Clinical Research and Study.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from ADNI. Restrictions apply to the availability of these data, which were used under license for this study. Data are available from <https://adni.loni.usc.edu/> with the permission of ADNI.

ORCID

Sung Hee Park  <https://orcid.org/0000-0002-4754-2611>

Liang Li  <https://orcid.org/0000-0001-5453-3839>

Lei Liu  <https://orcid.org/0000-0003-1844-338X>

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article.

References

- Akaike, H. (1974). A new look at the statistical model identification. *IEEE transactions on automatic control*, 19(6), 716–723.
- Alzheimer's Association. (2021). Special report—race, ethnicity and alzheimer's in america. *Alzheimers and Dementia*, 17(3).
- Anderson, J. R., Cain, K. C., & Gelber, R. D. (1983). Analysis of survival by tumor response. *J Clin Oncol*, 1(11), 710–719.
- Bi, X., Tang, X., Yuan, Y., Zhang, Y., & Qu, A. (2021). Tensors in statistics. *Annual review of statistics and its application*, 8, 345–368.
- Brier, G. W. (1950). Verification of forecasts expressed in terms of probability. *Monthly weather review*, 78(1), 1–3.
- Chang, S.-M., Yang, M., Lu, W., Huang, Y.-J., Huang, Y., Hung, H., ... Tzeng, J.-Y. (2021). Gene-set integrative analysis of multi-omics data using tensor-based association test. *Bioinformatics*, 37(16), 2259–2265.
- Cook, R. D., & Zhang, X. (2015). Foundations for envelope models and methods. *Journal of the American Statistical Association*, 110(510), 599–611.
- Cox, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society: Series B (Methodological)*, 34(2), 187–202.
- Dafni, U. (2011). Landmark analysis at the 25-year landmark point. *Circulation: Cardiovascular Quality and Outcomes*, 4(3), 363–371.
- Gauthier, S., Rosa-Neto, P., Morais, J. A., & Webster, C. (2021). World alzheimer report 2021: Journey through the diagnosis of dementia. *Alzheimer's Disease International*, 2022, 30.

- Gerds, T. A., & Schumacher, M. (2006). Consistent estimation of the expected brier score in general survival models with right-censored event times. *Biometrical Journal*, 48(6), 1029–1040.
- Graf, E., Schmoor, C., Sauerbrei, W., & Schumacher, M. (1999). Assessment and comparison of prognostic classification schemes for survival data. *Statistics in medicine*, 18(17-18), 2529–2545.
- Harrell, F. E., Califf, R. M., Pryor, D. B., Lee, K. L., & Rosati, R. A. (1982). Evaluating the yield of medical tests. *Jama*, 247(18), 2543–2546.
- Håstad, J. (1989). Tensor rank is np-complete. In *Automata, languages and programming: 16th international colloquium stresa, italy, july 11–15, 1989 proceedings 16* (pp. 451–460).
- Johansen, S. (1983). An extension of cox's regression model. *International Statistical Review/Revue Internationale de Statistique*, 165–174.
- Kolda, T. G., & Bader, B. W. (2009). Tensor decompositions and applications. *SIAM review*, 51(3), 455–500.
- Li, L., Luo, S., Hu, B., & Greene, T. (2017). Dynamic prediction of renal failure using longitudinal biomarkers in a cohort study of chronic kidney disease. *Statistics in biosciences*, 9, 357–378.
- Li, L., & Zhang, X. (2017). Parsimonious tensor response regression. *Journal of the American Statistical Association*, 112(519), 1131–1146.
- Li, X., Xu, D., Zhou, H., & Li, L. (2018). Tucker tensor regression and neuroimaging analysis. *Statistics in Biosciences*, 10(3), 520–545.
- Liang, H., & Zou, G. (2008). Improved aic selection strategy for survival analysis. *Computational statistics & data analysis*, 52(5), 2538–2548.
- Lin, D. (2007). On the breslow estimator. *Lifetime data analysis*, 13, 471–480.
- Lock, E. F. (2018). Tensor-on-tensor regression. *Journal of Computational and Graphical Statistics*, 27(3), 638–647.
- Lyu, R., Vannucci, M., & Kundu, S. (2024). Bayesian tensor modeling for image-based classification of alzheimer's disease. *Neuroinformatics*, 1–19.
- Mai, Q., & Zhang, X. (2023). Statistical methods for tensor data analysis. In *Springer handbook of engineering statistics* (pp. 817–829). Springer.
- Mirabnahrzazam, G., Ma, D., Beaulac, C., Lee, S., Popuri, K., Lee, H., ... others (2023). Predicting time-to-conversion for dementia of alzheimer's type using multi-modal deep survival analysis. *Neurobiology of aging*, 121, 139–156.
- Miranda, M. F., Zhu, H., Ibrahim, J. G., Initiative, A. D. N., et al. (2018). Tprm: Tensor partition regression models with applications in imaging biomarker detection. *The annals of applied statistics*, 12(3), 1422.
- Murtaugh, P. A., Dickson, R. E., Van Dam, G. M., Malinchoc, M., Grambsch, P. M., Langworthy, A. L., & Gips, C. H. (1994). Primary biliary cirrhosis: prediction of short-term survival based on repeated patient visits. *Hepatology*, 20(1), 126–134.
- Nakagawa, T., Ishida, M., Naito, J., Nagai, A., Yamaguchi, S., Onoda, K., & Initiative, A. D. N. (2020). Prediction of conversion to alzheimer's disease using deep survival analysis of mri images. *Brain communications*, 2(1), fcaa057.
- Nygård, S., Borgan, Ø., Lingjærde, O. C., & Størvold, H. L. (2008). Partial least squares cox regression for genome-wide data. *Lifetime Data Analysis*, 14, 179–195.
- Orozco-Sanchez, J., Trevino, V., Martinez-Ledesma, E., Farber, J., & Tamez-Peña, J. (2019). Exploring survival models associated with mci to ad conversion: a machine learning approach. *BioRxiv*, 836510.
- Rao, C. R., & Mitra, S. K. (1972). Generalized inverse of a matrix and its applications. In *Proceedings of the sixth berkeley symposium on mathematical statistics and probability, volume 1: Theory of statistics* (Vol. 6, pp. 601–621).
- Schwarz, G. (1978). Estimating the dimension of a model. *The annals of statistics*, 461–464.
- Sidiropoulos, N. D., & Bro, R. (2000). On the uniqueness of multilinear decomposition of n-way arrays. *Journal of Chemometrics: A Journal of the Chemometrics Society*, 14(3), 229–239.
- Spencer, D., Guhaniyogi, R., Shinohara, R., & Prado, R. (2022). Bayesian tensor regression using the tucker decomposition for sparse spatial modeling. *arXiv preprint arXiv:2203.04733*.
- Steyerberg, E. (2009). Applications of prediction models. In *Clinical prediction models: A practical approach to development, validation, and updating* (pp. 11–31). Springer New York.
- Sun, W. W., Hao, B., & Li, L. (2021). Tensors in modern statistical learning. *Wiley StatsRef: Statistics Reference Online*, 1–25.
- Therneau, T. M., Grambsch, P. M., Therneau, T. M., & Grambsch, P. M. (2000). *The cox model*. Springer.
- Van Houwelingen, H., & Putter, H. (2011). *Dynamic prediction in clinical survival analysis*. CRC Press.
- Volinsky, C. T., & Raftery, A. E. (2000). Bayesian information criterion for censored survival models. *Biometrics*, 56(1), 256–262.

- Weiner, M. W., Aisen, P. S., Jack Jr, C. R., Jagust, W. J., Trojanowski, J. Q., Shaw, L., ... others (2010). The alzheimer's disease neuroimaging initiative: progress report and future plans. *Alzheimer's & Dementia*, 6(3), 202–211.
- Yao, Y., Li, L., Astor, B., Yang, W., & Greene, T. (2023). Predicting the risk of a clinical event using longitudinal data: The generalized landmark analysis. *BMC Medical Research Methodology*, 23(1), 5.
- Zhang, X., & Li, L. (2017). Tensor envelope partial least-squares regression. *Technometrics*, 59(4), 426–436.
- Zheng, Y., & Heagerty, P. J. (2005). Partly conditional survival models for longitudinal data. *Biometrics*, 61(2), 379–391.
- Zhou, H., Li, L., & Zhu, H. (2013). Tensor regression with applications in neuroimaging data analysis. *Journal of the American Statistical Association*, 108(502), 540–552.

AUTHOR BIOGRAPHY

Sung Hee Park is a Postdoctoral Researcher at Informatics, Data Science and Biostatistics (I²DB) at Washington University in St. Louis. He received his B.S. and M.S. degrees in Statistics from Dongguk University in 2014 and 2016, respectively, and his Ph.D. in Statistics from Florida State University in 2023. His research interests include medical imaging, longitudinal data analysis, canonical correlation analysis, tensor data analysis, and dimension reduction.

Ruiwen Zhou is a Postdoctoral Researcher at I²DB at Washington University in St. Louis. Her research interests include survival analysis, joint modeling, missing data, and deep learning methods for time-to-event data prediction.

Xin (Henry) Zhang is a Professor in the Department of Statistics at the Florida State University. He received his B.S. degree in Applied Physics from the University of Science and Technology of China in 2008 and his Ph.D. in Statistics from the University of Minnesota in 2014. His current research interests include tensor analysis, multivariate and computational statistics, medical imaging, dimension reduction, and envelope models. He received the Outstanding Young Researcher Award from the International Chinese Statistical Association in 2020 and the 2021 Developing Scholar Award at FSU. He is currently serving as Associate Editor of *Biometrics* and *Statistica Sinica*.

Liang Li is a Professor at the Department of Biostatistics, University of Texas MD Anderson Cancer Center. He received his B.S. degree in Genetics and Cell biology at Peking University and his Ph.D. in Statistics from University of Wisconsin Madison. His research interests contains dynamic prediction models, joint modeling of longitudinal and survival data, propensity score analysis, observational studies, electronic health records, medical costs modeling and early phase biomarker studies. He collaborates with clinicians in various fields, such as, chronic disease epidemiology, pancreatic cancer, hepatocellular carcinoma, health services research, population health and behavioral sciences.

Lei Liu is a Professor at I²DB at Washington University in St. Louis and a fellow of the American Statistical Association. He received his B.S. and M.S. degrees in Engineering from Zhejiang University and his Ph.D. in Biostatistics from the University of Michigan. He has diverse research interests in biostatistical and data science methods, including survival analysis, longitudinal data analysis, spline regression, personalized medicine, machine learning, and artificial intelligence. His research is focused on analyzing high-dimensional omics (epigenetics and microbiome) data, medical cost data, and joint models of multi-outcome data. He collaborates with clinicians in various medical fields, for example, cancer, cardiovascular, addiction, ophthalmology, nephrology, infectious disease, asthma, and diabetes.

□

APPENDIX

A MATRIX NORMAL DISTRIBUTION

Let $\mathbf{X}$ be a $p_1 \times p_2$ random matrix with $\mathbf{X} \sim \mathcal{MN}(\mathbf{M}, \Sigma_1, \Sigma_2)$, where $\mathbf{M}$ is the mean matrix and Σ_1, Σ_2 are positive-definite covariance matrices. The probability density function of the random matrix $\mathbf{X}$ is as follows:

$$p(\mathbf{X} | \mathbf{M}, \Sigma_1, \Sigma_2) = \frac{1}{(2\pi)^{\frac{p_1 p_2}{2}} \det(\Sigma_1)^{\frac{p_1}{2}} \det(\Sigma_2)^{\frac{p_2}{2}}} \exp \left\{ -\frac{1}{2} \text{trace} \left(\Sigma_1^{-1} (\mathbf{X} - \mathbf{M}) \Sigma_2^{-1} (\mathbf{X} - \mathbf{M})^\top \right) \right\},$$

where $\det(\cdot)$ denotes the determinant and $\text{trace}(\cdot)$ denotes the trace of the matrix. The matrix normal distribution is connected to the multivariate normal distribution by the following relationship:

$$\mathbf{X} \sim \mathcal{MN}(\mathbf{M}, \Sigma_1, \Sigma_2) \text{ if and only if } \text{vec}(\mathbf{X}) \sim \mathcal{N}_{p_1 p_2}(\text{vec}(\mathbf{M}), \Sigma_2 \otimes \Sigma_1),$$

where $\otimes$ denotes the Kronecker product.

B ASSESSMENTS FOR PREDICTION

B.1 C-index

The C-index is defined as

$$\text{C-index} = \frac{\sum_{i,j} \mathbf{I}_{t_i < t_j} \cdot \mathbf{I}_{\zeta_i > \zeta_j} \delta_i}{\sum_{i,j} \mathbf{I}_{t_i < t_j} \cdot \delta_i}, \quad i, j = 1, \dots, n,$$

where ζ_i is the predicted risk score for subject i , calculated as $\zeta_i = \eta^\top \mathbf{z}_i + \langle \mathbf{B}, \mathbf{X}_i \rangle$.

B.2 Integrated IPCW Brier score

A Brier score (Brier, 1950) is a metric used to evaluate the accuracy of probabilistic predictions. For a binary outcome, the Brier score is defined as

$$\text{Brier} = \frac{1}{n} \sum_{i=1}^n (f_i - o_i)^2,$$

where n is the number of observations for prediction, f_i is the predicted probability of the event occurring for the i -th observation, and o_i is the actual outcome for instance i where $o_i = 1$ if the event occurred and $o_i = 0$ if it did not.

The Brier score ranges from 0 to 1, where 0 indicates perfect accuracy, meaning the predictions exactly match the actual outcomes, and 1 represents the worst possible accuracy, reflecting completely incorrect predictions. The Brier score measures the probabilistic difference between predicted and actual outcomes, providing a comprehensive assessment beyond just discrimination. The Brier score is typically examined as follows:

$$0 \sim 0.25 : \text{good}, \quad 0.25 \sim 0.5 : \text{moderate}, \quad 0.5 \sim 1 : \text{poor}.$$

The inverse probability of censoring weighting (IPCW) Brier score (Graf et al., 1999; Gerds & Schumacher, 2006) for survival analysis extends the traditional Brier score to handle censored data. For a survival model with right-censored data, the IPCW Brier score at a specific time t is defined as

$$\text{Brier}_{\text{IPCW}}(t) = \frac{1}{n} \sum_{i=1}^n w_i(t) \left\{ \widehat{S}(t | \mathbf{X}_i, \mathbf{z}_i) - \mathbf{I}(t_i > t) \right\}^2,$$

where n is the number of individuals, $\widehat{S}(t | \mathbf{X}_i, \mathbf{z}_i)$ is the predicted survival probabilities at time t for individual i given matrix covariates $\mathbf{X}_i$, t_i is the observed time for individual i (either the event time or the censoring time), $\mathbf{z}_i$ is a time-independent covariate vector, $\mathbf{I}(t_i \leq t)$ is an indicator function that equals 1 if the event occurred before or at time t and 0 otherwise, and $w_i(t)$ is the inverse probability of censoring weight for individual i at time t .

The weight $w_i(t)$ adjusts for the bias introduced by censored observations by weighting the contribution of each observation according to the inverse probability of remaining uncensored up to time t . The weight $w_i(t)$ is defined as

$$w_i(t) = \frac{\delta_i}{G(t | \mathbf{X}_i, \mathbf{Z}_i)},$$

where $G(t | \mathbf{X}_i, \mathbf{Z}_i)$ is the estimated probability of being uncensored at time t for individual i . The censoring distribution $G(t)$ can be estimated using the Kaplan-Meier estimator applied to the censoring times. The integrated IPCW Brier Score (IBS) provides a single summary measure of the IPCW Brier Score over a range of times. The IBS from landmark time s to τ year prediction is defined as follows:

$$\text{IBS} = \int_s^{s+\tau} \text{Brier}_{\text{IPCW}}(t) dt,$$

where τ is the time horizon for evaluating the predictions.

C VARIABLE DESCRIPTION OF ADNI DATA

Name	Description
PTGENDER	Sex (male, female)
AGE	Baseline age for participants
DX_bl	Baseline diagnosis (CN,SMC,EMCI,LMCI)
PTEDUCAT	Education (4-20)
APOE4	Number of APOE ϵ 4 alleles
ADAS11	Alzheimer's Disease Assessment Scale (ADAS) 11
ADAS13	ADAS 13 (including delayed word recall and number cancelation)
ADASQ4	ADAS Delayed Word Recall
Entorhinal	University of California, San Francisco (UCSF) Entorhinal
Fusiform	UCSF Fusiform
Hippocampus	UCSF Hippocampus
ICV	UCSF ICV
MidTemp	UCSF Middle temporal gyrus
Ventricles	UCSF Ventricles
WholeBrain	UCSF WholeBrain
RAVLT.forgetting	Rey Auditory Verbal Learning Test (RAVLT) forgetting
RAVLT.immediate	RAVLT immediate
RAVLT.learning	RAVLT Learning
RAVLT.perc.forgetting	RAVLT percent forgetting
TRABSCOR	Trail-making test B (Trails B) score
mPACCtrailsB	ADNI-modified Preclinical Alzheimer's Cognitive Composite (PACC) with Trails B
mPACCdigit	ADNI-modified PACC with digit symbol substitution
LDELTOTAL	Logical memory - delayed recall
CDRSB	Clinical Dementia Rating scale Sum of Boxes
FAQ	Functional Assessment Questionnaire (0-30)
MMSE	Mini-Mental State Examination (MMSE) score

TABLE C1 List of covariates used for the ADNI data analysis. The gray rows indicate baseline covariates used for analysis.