FLCRM: Functional Linear Cox Regression Model

Dehan Kong,[1] Joseph G. Ibrahim,[2] Eunjee Lee,[3] and Hongtu Zhu [iD] [4,*]

[1]Department of Statistical Sciences, University of Toronto, Ontario, Canada
[2]Department of Biostatistics, University of North Carolina at Chapel Hill, North Carolina, U.S.A.
[3]Department of Biostatistics, University of Michigan, Michigan, U.S.A.
[4]Department of Biostatistics, The University of Texas MD Anderson Cancer Center, Texas, U.S.A.
[*]*email:* htzhu@email.unc.edu

SUMMARY. We consider a functional linear Cox regression model for characterizing the association between time-to-event data and a set of functional and scalar predictors. The functional linear Cox regression model incorporates a functional principal component analysis for modeling the functional predictors and a high-dimensional Cox regression model to characterize the joint effects of both functional and scalar predictors on the time-to-event data. We develop an algorithm to calculate the maximum approximate partial likelihood estimates of unknown finite and infinite dimensional parameters. We also systematically investigate the rate of convergence of the maximum approximate partial likelihood estimates and a score test statistic for testing the nullity of the slope function associated with the functional predictors. We demonstrate our estimation and testing procedures by using simulations and the analysis of the Alzheimer's Disease Neuroimaging Initiative (ADNI) data. Our real data analyses show that high-dimensional hippocampus surface data may be an important marker for predicting time to conversion to Alzheimer's disease. Data used in the preparation of this article were obtained from the ADNI database (adni.loni.usc.edu).

KEY WORDS: Cox regression; Functional predictor; Functional principal component analysis; Score test.

1. Introduction

Roughly more than 5 million Americans are suffering from memory loss and dementia caused by Alzheimer's Disease (AD) and it costs the nation approximately 203 billion dollars in medical expenses solely in 2013. An earlier and more accurate diagnosis of AD is considered to be an important goal for researchers because therapeutic intervention is more likely to be more beneficial during the early development of the disease. This led to the development of Mild Cognitive Impairment (MCI), which is a transitional stage between normal aging and the development of AD (Petersen, 2004). MCI is characterized by insidious onset and gradual progression, and commonly arises as a result of underlying neurodegenerative pathology. There is a substantial interest in delineating a set of biomarkers that provide evidence of such neurodegenerative pathology in living individuals, with the goal of specifying the likelihood that the pathophysiological process is due to Alzheimer's disease (MCI-AD) and will lead to dementia within a few years. Accordingly, increasing attention has been devoted to investigate the utility of various imaging, genetic, clinical, behavioral, and fluid data to predict the conversion from MCI to AD (Risacher et al., 2009).

The development of functional linear Cox regression model (FLCRM) is motivated by addressing a critical question: "how do we accurately predict the time to conversion in individuals who harbor AD pathology, as well as assess the predictive role of surface morphology?" There is a long-term interest in answering this question. Li et al. (2013) employed data from 139 MCI subjects in ADNI to evaluate the predictive

power of brain volume, ventricular volume, hippocampus volume, APOE status, cerebrospinal fluid (CSF) biomarkers, and behavioral scores. Da et al. (2014) used 381 MCI subjects from ADNI to evaluate several biomarkers for predicting MCI to AD conversion including spatial patterns of brain atrophy, ADAS-Cog score, APOE genotype, and cerebrospinal fluid (CSF) biomarkers. To the best of our knowledge, no prior study has examined the role of neuroimaging data in predicting time to conversion from MCI to AD, while adjusting for low-dimensional behavioral and clinical measures.

The aim of this article is to consider a FLCRM to examine the relationship of the survival distribution to a set of functional and scalar predictors in large-scale biomedical studies. Let the random variables T, C, and $\widetilde{T} = \min\{T, C\}$ be, respectively, the failure time, censoring time, and observed time, and $f(t)$ and $S(t)$, respectively, denote the probability density function and the survival function of T. The Cox regression model (Cox, 1972) has been widely used to model survival time T as a function of p predictors, denoted by $Z = (z_1, \ldots, z_p)^{\mathrm{T}}$. The hazard function for the Cox regression model has the form

$$h(t) = f(t)/S(t) = h_0(t) \exp(Z^{\mathrm{T}} \gamma), \tag{1.1}$$

where $h_0(t)$ is a completely unspecified baseline hazard function and $\gamma = (\gamma_1, \ldots, \gamma_p)^{\mathrm{T}}$. Although the Cox regression model and its various extensions have been widely investigated for a small number of predictors (Cox, 1975; Ibrahim et al., 2001; Kalbfleisch and Prentice, 2002), there is a great

interest in developing accurate risk prediction models with high-dimensional genomic and clinical data to predict censored survival outcomes (Cai et al., 2011; Li and Ma, 2013).

Existing survival models for high-dimensional genomic data, however, suffer from a major limitation of incorporating infinite-dimensional imaging data to predict survival outcomes. Compared with genetic and clinical data, a key unique feature of imaging data is that they are high dimensional and intrinsically continuous functions measured at a large number of grid points. The effect of imaging data on survival outcomes is often non-sparse, which makes inference notoriously difficult based on existing regularization methods (Xu, 2012; Huang et al., 2013). Therefore, it is imperative to use some dimension reduction methods to extract and select "low-dimensional" important features, while eliminating redundant features (Johnstone and Lu, 2009).

Functional linear regression has become a standard method in functional data analysis for incorporating functional predictors. Most functional linear regression models focus on modeling the relationship between a functional or continuous response, Y, and a one-dimensional functional predictor $X(s)$, in which s varies in a compact set $\mathcal{S}$. Functional linear regression usually assumes

$$Y = \int_{\mathcal{S}} X(s)\beta(s)ds + \epsilon, \tag{1.2}$$

where ϵ is a noise term independent of $X(s)$ and $\beta(s)$ is an unknown function of interest. Two popular estimation methods for $\beta(s)$ include the functional principal component analysis (FPCA) and methods of penalization. A few examples include Hall and Horowitz (2007); Reiss and Ogden (2007, 2010); Morris (2015) and the references therein. However, very little has been done on modeling survival outcomes and scalar and functional predictors.

We consider a FLCRM that incorporates (1.1) and (1.2) for modeling the relationship between survival outcomes and a set of finite and infinite dimensional predictors. Specifically, the hazard function for the FLCRM has the form

$$\log\{h(t)\} = \log\{h_0(t)\} + \int_{\mathcal{S}} X(s)\beta(s)ds + Z^{\mathrm{T}}\gamma. \tag{1.3}$$

At the time of submission, we are aware of three recent articles on the development of various estimation methods for model (1.3). Gellar et al. (2015) and Qu et al. (2016) proposed to maximize penalized partial likelihood functions for model (1.3), whereas Lee et al. (2015) developed a Bayesian framework for the same model. In particular, Gellar et al. (2015) combined penalized signal regression with methods developed for mixed effects proportional hazards models under penalized B-spline framework, and Qu et al. (2016) estimated the model under the reproducing kernel Hilbert space framework.

Compared to the existing literature, we make several new contributions. To deal with the nonparametric function $\beta(s)$, we employ a FPCA method and then approximate $\beta(s)$ by the eigen-functions of the covariance operator of the functional predictor $X(s)$. Such a FPCA method has been widely used in various functional linear regression models (Yao et al., 2005; Hall and Hosseini-Nasab, 2006). In particular, we use the first several functional principal components (FPCs) to represent the infinite dimensional process $X(s)$, and the number of FPCs is treated as a tuning parameter diverging with the sample size, which distinguishes our work from the ones that fix the number of FPCs (Li, Wang, and Carroll, 2010). Subsequently, our FLCRM reduces to a Cox regression model with a growing number of predictors. We also examine both estimation and testing for FLCRM, and none of the recent articles on the development of various estimation methods for model (1.3) (Gellar et al., 2015; Lee et al., 2015; Qu et al., 2016) consider the testing problem. We further establish their associated asymptotic properties, and study the asymptotic theories when the number of scalar parameters diverges at a polynomial rate of sample size. In addition, we shed lights on how to choose the number of FPCs for both estimation and testing procedures. It turns out that we need to use different criteria for estimation and testing procedures.

2. Functional Linear Cox Regression Models

2.1. *Model Setup*

Let $\mathcal{S}$ be a compact set of $\mathbb{R}$ and $X_i(\cdot)$ be the true trajectory of the ith individual. We observe the curve on grid points $\{s_{im} \in \mathcal{S}, 1 \le m \le M_i\}$ with some measurement errors such that $W_{im} = X_i(s_{im}) + \epsilon_{im}$, where the ϵ_{im}'s are independent and identically distributed random variables with zero mean and variance σ_e^2. We consider a random sample of n subjects and observe $(\{W_{1m}, 1 \le m \le M_1\}, Z_1, \delta_1, \tilde{T}_1), \ldots, (\{W_{nm}, 1 \le m \le M_n\}, Z_n, \delta_n, \tilde{T}_n)$, where $\delta_i = 1(T_i \le C_i)$, which equals 1 if the observed event is a failure and 0 otherwise. In neuroimaging studies, W_{im} and Z_i, respectively, denote the imaging measure at s_{im} and scalar predictors such as age, gender, or candidate genetic marker.

For model (1.3), we obtain the hazard function of the i-th subject under Cox regression as

$$h_i(t) = h_0(t) \exp\left\{ \sum_{k=1}^{p} z_{ik}\gamma_k + \int_{\mathcal{S}} X_i(s)\beta(s)ds \right\}, \tag{2.1}$$

where $Z_i = (z_{i1}, \ldots, z_{ip})^{\mathrm{T}}$, and $t \in [0, \tau]$ for some finite $\tau > 0$. The FLCRM is determined by the unknown coefficient function $\beta(\cdot)$, the unknown parameter vector $\gamma = (\gamma_1, \ldots, \gamma_p)^{\mathrm{T}}$, and the baseline hazard function $h_0(\cdot)$.

The second component of FLCRM is the FPCA model of both $X(s)$ and $\beta(s)$. Let $\mu(s)$ and $K(s, s')$ be, respectively, the mean and covariance functions of the stochastic process $\{X(s) : s \in S\}$ determined by the functional predictors. Let the spectral decomposition of the covariance function $K(s, s') = \sum_{j=1}^{\infty} \lambda_j \phi_j(s)\phi_j(s')$, where $\{\lambda_j, j \ge 1\}$ are the eigenvalues in decreasing order with $\sum_{j=1}^{\infty} \lambda_j < \infty$ and the ϕ_j's are the corresponding eigenfunctions. Thus, the i-th trajectory $X_i(\cdot)$ can be represented by using the Karhunen–Loève expansion as

$$X_i(s) = \mu(s) + \sum_{j=1}^{\infty} \xi_{ij}\phi_j(s).$$

The i-th observed trajectory is defined as $W_i(s) = X_i(s) + \epsilon_i(s)$, where $\epsilon_i(s)$ is measurement error with mean zero and variance $\sigma^2(s)$ at s and is independent of $\epsilon_i(s')$ for $s \neq s'$. Moreover, $\xi_{ij} = \int \{X_i(s) - \mu(s)\}\phi_j(s)ds$ is the FPC score and has mean zero with $\text{cov}(\xi_{ij}, \xi_{ik}) = \lambda_j 1(j = k)$. We also consider the FPCA model of $\beta(s)$. Based on the basis $\{\phi_j(s) : 1 \leq j < \infty\}$, it is assumed that the projection of $\beta(\cdot)$ onto the span of K is identifiable, we can expand $\beta(s)$ as $\beta(s) = \sum_{j=1}^{\infty} \phi_j(s)\beta_j$. Thus, we can calculate $\int_{\mathcal{S}} X_i(s)\beta(s)ds$ as $\int_{\mathcal{S}} \mu(s)\beta(s)ds + \sum_{j=1}^{\infty} \xi_{ij}\beta_j$, where $\beta_j = \int_{\mathcal{S}} \beta(s)\phi_j(s)ds$.

Based on the FPCA model, the hazard function of FLCRM can be rewritten as

$$h_i(t) = h_0^*(t)\exp\left(\sum_{k=1}^{p} z_{ik}\gamma_k + \sum_{j=1}^{\infty} \xi_{ij}\beta_j\right), \qquad (2.2)$$

where $h_0^*(t) = h_0(t)\exp\{\int_{\mathcal{S}} \mu(s)\beta(s)ds\}$. Since the number of predictors in (2.2) is infinite, we propose to approximate $h_i(t)$ by truncating the number of FPC scores, denoted as r_n, which increases asymptotically as $n \to \infty$. Thus, FLCRM reduces to a Cox regression model with high-dimensional predictors as

$$h_i(t) \approx h_0^*(t)\exp\left(\sum_{k=1}^{p} z_{ik}\gamma_k + \sum_{j=1}^{r_n} \xi_{ij}\beta_j\right). \qquad (2.3)$$

The approximation (2.3) depends on whether the slope function $\beta(\cdot)$ is efficiently represented in terms of the leading eigenfunctions or equivalently $\sum_{r_n+1}^{\infty} |\xi_{ij}\beta_j| \approx 0$. We will discuss how to choose r_n in Section 2.4.

2.2. Estimation Procedure

We develop a three-step estimation procedure as follows.

- Step (I): We use the local linear regression technique to smooth all individual functions $\{X_i(s_m) : m = 1, \ldots, M\}$. It leads to a smoothed estimate of $X_i(s)$, denoted by $\widehat{X}_i(s)$.
- Step (II): Estimate $K(s, s')$ and its eigenvalues and eigenfunctions and then calculate the estimated FPC scores of all subjects.
- Step (III): Calculate the maximum approximate partial likelihood estimate of $\eta = (\beta_n^{\text{T}}, \gamma^{\text{T}})^{\text{T}}$, where $\beta_n = (\beta_1, \ldots, \beta_{r_n})^{\text{T}}$. Finally, we use the Nelson–Aalen method to estimate the baseline hazard function.

Step (I) is to estimate $\widehat{X}_i(s)$ for all $s \in \mathcal{S}$ and $i = 1, \ldots, n$ by using the well-known local linear regression technique (Fan and Gijbels, 1996). Let $\partial_s X_i(s)$ be $\partial X_i(s)/\partial s$ and $K_{loc}(\cdot)$ be a kernel function. For each i, we calculate $\widehat{X}_i(s) = (1, 0)\widehat{C}_i(s)$ by minimizing the weighted least squares function as

$$\widehat{C}_i(s) = \text{argmin}_{C_i(s)} \sum_{m=1}^{M_i} \left\{ W_{im} - C_i(s)^{\text{T}} Z_h(s_m - s) \right\}^2 K_h(s_m - s),$$

where $C_i(s) = (X_i(s), h\{\partial_s X_i(s)\}^{\text{T}})^{\text{T}}$, $Z_h(s_m - s) = (1, (s_m - s)/h)^{\text{T}}$ are two dimensional vectors with each component a nonparametric function, and $K_h(s_m - s) = K_{loc}\{(s_m - s)/h\}$ is the

rescaled kernel function with a bandwidth h. We pool the data from all n subjects and select the optimal bandwidth $\tilde{h}$ by minimizing the generalized cross-validation score (Zhang and Chen, 2007).

In Step (II), we calculate $\widehat{\mu}(s) = n^{-1}\sum_{i=1}^{n} \widehat{X}_i(s)$ and the sample covariance function of $\{\widehat{X}(s) : s \in \mathcal{S}\}$ given by

$$\widehat{K}(s, s') = n^{-1}\sum_{i=1}^{n} \left\{\widehat{X}_i(s) - \widehat{\mu}(s)\right\} \left\{\widehat{X}_i(s') - \widehat{\mu}(s')\right\}$$

$$= \sum_{j=1}^{\infty} \widehat{\lambda}_j \widehat{\phi}_j(s)\widehat{\phi}_j(s'),$$

where $\{\widehat{\lambda}_j, j \geq 1\}$ are the estimated eigenvalues in decreasing order with $\sum_{j=1}^{\infty} \widehat{\lambda}_j < \infty$ and $\{\widehat{\phi}_j(\cdot), j \geq 1\}$ are the corresponding estimated eigenfunctions. Finally, the estimated eigenscores can be calculated by $\widehat{\xi}_{ij} = \int_{\mathcal{S}}\{X_i(s) - \widehat{\mu}(s)\}\widehat{\phi}_j(s)ds$ for all $j \leq r_n$, where choice of the truncation number r_n would be discussed in Section 2.4.

In Step (III), we plug the first r_n estimated FPCs scores into (2.3) to obtain a Cox regression model with $r_n + p$ predictors. Let $N_i(t) = 1(\tilde{T}_i \leq t, \delta_i = 1)$, $\overline{N}(t) = \sum_{i=1}^{n} N_i(t)$, and $R(t) = \{j : \tilde{T}_j \geq t\}$ be the set of subjects who are at risk and uncensored prior to time t. For $i = 1, \ldots, n$, we define $Y_i(t) = 1(\tilde{T}_i \geq t) = 1(i \in R(t))$. In this article, we assume that the observed failure times are distinct. If there are ties, one may use Efron's approximation (Efron, 1977), for example. Thus, the log-approximate partial likelihood function of η, denoted as $Q(\eta)$, is given by

$$Q(\eta) = \sum_{i=1}^{n} \int_0^{\tau} \widehat{w}_i^{\text{T}} \eta \, dN_i(t) - \int_0^{\tau} \log\left\{\sum_{i=1}^{n} Y_i(t)\exp(\widehat{w}_i^{\text{T}}\eta)\right\} d\overline{N}(t),$$

$$(2.4)$$

where $\widehat{w}_i = (\widehat{\xi}_{i1}, \ldots, \widehat{\xi}_{ir_n}, z_{i1}, \ldots, z_{ip})^{\text{T}}$. We use the Newton–Raphson algorithm as implemented in the R function "`coxph()`" under the `survival` package to calculate the maximum approximate partial likelihood estimate, denoted as $\widehat{\eta} = (\widehat{\beta}_n^{\text{T}}, \widehat{\gamma}^{\text{T}})^{\text{T}}$, which maximizes $Q(\eta)$, and their standard errors. Subsequently, we use the Nelson–Aalen method to estimate the cumulative baseline hazard function, $H_0(t) = \int_0^{\text{T}} h_0(u)du$, by $\widehat{H}_0(t) = \sum_{i:\tilde{T}_i \leq t}[\delta_i/\{\sum_{j \in R(\tilde{T}_i)} \exp(\widehat{w}_i^{\text{T}}\widehat{\eta})\}]$. After we get $\{\widehat{\beta}_j, 1 \leq j \leq r_n\}$, we can calculate the estimated coefficient function by $\widehat{\beta}(s) = \sum_{j=1}^{r_n} \widehat{\beta}_j\widehat{\phi}_j(s)$. A limitation of our method is that it is difficult to obtain the confidence bands of $\widehat{\beta}(\cdot)$, which is an interesting topic for future research.

2.3. Testing Procedure

In real applications, evidence for the association between a functional predictor and the survival outcome is as valuable as, if not more than, estimation of the actual effect size. For example, in the ADNI dataset, it would be of great interest to test whether the hippocampus shape information is an important biomarker for predicting time to conversion to Alzheimer disease in patients with MCI. We propose a score test to statistically test the null effect of a functional predictor

on the time-to-event data with the presence of other scalar predictors. Specifically, we are interested in testing

$$H_0 : \beta(s) = 0 \quad \text{versus} \quad H_a : \beta(s) \neq 0. \quad (2.5)$$

The testing problem under the functional linear regression context has been studied in the literature (Cardot et al., 2003; Kong et al., 2016).

Since $\beta(s)$ is an infinite dimensional parameter, we instead truncate the number of basis functions for $\beta(s)$ and then test

$$H_0^* : \beta_1 = \beta_2 = \ldots = \beta_{r_n} = 0 \quad \text{versus}$$
$$H_a^* : \beta_j \neq 0 \text{ for at least one } j, 1 \leq j \leq r_n. \quad (2.6)$$

Our test is based on the score test statistic for FLCRM (2.3). We first derive the score function and the information matrix of η for the approximate partial likelihood function (2.4) as:

$$S(\eta) = \frac{\partial Q(\eta)}{\partial \eta} = \sum_{i=1}^{n} \int_0^{\tau} \widehat{w}_i dN_i(t)$$
$$- \int_0^{\tau} \frac{\sum_{i=1}^{n} \widehat{w}_i Y_i(t) \exp\{\widehat{w}_i^{\mathrm{T}} \eta\}}{\sum_{i=1}^{n} Y_i(t) \exp\{\widehat{w}_i^{\mathrm{T}} \eta\}} d\overline{N}(t),$$

$$I(\eta) = -\frac{\partial S(\eta)}{\partial \eta} = \int_0^{\tau} \left[\frac{\sum_{i=1}^{n} \widehat{w}_i^{\otimes 2} Y_i(t) \exp(\widehat{w}_i^{\mathrm{T}} \eta)}{\sum_{i=1}^{n} Y_i(t) \exp(\widehat{w}_i^{\mathrm{T}} \eta)} \right.$$
$$\left. - \left\{ \frac{\sum_{i=1}^{n} \widehat{w}_i^{\mathrm{T}} Y_i(t) \exp(\widehat{w}_i^{\mathrm{T}} \eta)}{\sum_{i=1}^{n} Y_i(t) \exp(\widehat{w}_i^{\mathrm{T}} \eta)} \right\}^{\otimes 2} \right] d\overline{N}(t),$$

where $a^{\otimes 2} = aa^{\mathrm{T}}$. Denote $\widehat{\eta}_0 = (0^{\mathrm{T}}, \widehat{\gamma}_0^{\mathrm{T}})^{\mathrm{T}}$ as the maximum approximate partial likelihood estimate of η under H_0^*. We have

$$I(\widehat{\eta}_0) = \int_0^{\tau} \left[\frac{\sum_{i=1}^{n} \widehat{w}_i^{\otimes 2} Y_i(t) \exp(z_i^{\mathrm{T}} \widehat{\gamma}_0)}{\sum_{i=1}^{n} Y_i(t) \exp(z_i^{\mathrm{T}} \widehat{\gamma}_0)} \right.$$
$$\left. - \left\{ \frac{\sum_{i=1}^{n} \widehat{w}_i^{\mathrm{T}} Y_i(t) \exp(z_i^{\mathrm{T}} \widehat{\gamma}_0)}{\sum_{i=1}^{n} Y_i(t) \exp(z_i^{\mathrm{T}} \widehat{\gamma}_0)} \right\}^{\otimes 2} \right] d\overline{N}(t).$$

Assume the $(r_n + p) \times (r_n + p)$ dimensional information matrix $I(\widehat{\eta}_0)$ is invertible, the score test for testing H_0^*, denoted by T_S, is given by

$$T_S = S(\widehat{\eta}_0)^{\mathrm{T}} \{I(\widehat{\eta}_0)\}^{-1} S(\widehat{\eta}_0).$$

Similar to the traditional score statistic in the Cox regression model, we will show that $\chi_{r_n}^2$ provides a good approximation to the null distribution of the score statistic T_S.

2.4. *Choice of r_n*

We introduce two different methods of choosing r_n for our testing and estimation procedures. The first one used in

testing is to threshold the percentage of variance explained by the first r_n FPCs, denoted by $\mathrm{PV}(r_n) = \sum_{j=1}^{r_n} \lambda_j / \sum_{j=1}^{\infty} \lambda_j$. For instance, one may set the threshold values of $\mathrm{PV}(r_n)$ as 70, 85, or 95%. We have found in our simulation studies that the testing procedure is relatively robust under different threshold values, and thus we use $\mathrm{PV}(r_n) = 85\%$ throughout the article. The second one used in estimation is to choose an appropriate r_n for accurately estimating $\beta(t)$. We have found that changing r_n can have a relatively large effect on the accuracy of estimating $\beta(t)$. Generally, a smaller r_n would lead to a larger bias, whereas a larger r_n could lead to a larger variance. To achieve a balance between bias and variance, we use AIC to choose r_n (Yao et al., 2005). Let $\widehat{\eta}(r_n)$ be the estimate of η when the truncation integer is set as r_n, AIC is defined as $AIC(r_n) = 2r_n - 2\log\{Q(\hat{\eta}(r_n))\}$. Numerically, we can use the grid search method to select an optimal r_n that minimizes AIC. We will show in our simulations that AIC does not work well in testing, whereas thresholding the percentage of variance does not work well in estimation.

2.5. *Computational Efficiency and Asymptotic Theory*

Our method is easy to implement even for two- or higher dimensional images. Specifically, we only need to fit a typical cox regression model after we obtain the functional principal component (FPC) scores. Both steps are computationally fast. It takes 33 seconds to run the real data analysis by using our method. Among them, majority of time is spent on the construction of FPCA since we have 30,000 grid points on each image. We also systematically investigate the asymptotic properties of the maximum approximate partial likelihood estimator $\widehat{\eta}$ as well as the asymptotic null distribution of the score statistic T_S. We have included them in Section 4 of the Supplementary Material.

3. Simulations

3.1. *Estimation*

We simulated datasets from the FLCRM with the hazard function (2.1), in which four scalar predictors ($p = 4$) and one functional predictor were considered. Specifically, the failure time T was independently generated from an exponential distribution with parameter

$$h(t|X_i(s), Z_i) = h_0(t) \exp\left\{ \int_0^1 X_i(s)\beta_0(s)ds + Z_i^{\mathrm{T}}\gamma_0 \right\},$$

in which we set $h_0(t) = 1$, $\gamma_0 = (0.2, 0.2, 0.2, 0.2)^{\mathrm{T}}$, and $\beta_0(s) = 0.3[\sin(\pi s) - \cos(\pi s) + \sin(3\pi s/10) - \cos(3\pi s) + \sin(5\pi s)/9 - \cos(5\pi s)/9 + \sin(7\pi s)/16 - \cos(7\pi s)/16 + \sin(9\pi s)/25 - \cos(9\pi s)/25 + (2\pi)^{-1/2} \exp\{-2^{-1}(s - 0.5)^2\}]$ for $0 \leq s \leq 1$. The censoring time was then independently simulated from a uniform distribution $U(0, c_0)$, where c_0 was chosen to achieve a desired censoring rate of 10, 30, or 50%. We independently simulated $X_i(s) = u_{i1} + u_{i2}s + \sum_{j=1}^{10} [v_{ij1} \sin\{2(2j-1)\pi s\} + v_{ij2} \cos\{2(2j-1)\pi s\}]$, where $u_{i1} \sim N(0, 1)$, $u_{i2} \sim N(0, 1)$, and $v_{ij1}, v_{ij2} \sim N(0, 1/j^2)$. The scalar covariates Z_i were simulated from a multivariate normal distribution with zero mean and covariance matrix $\Sigma = (\rho^{|j-k|})_{1 \leq j,k \leq 4}$ with $\rho = 0.5$. We allow some correlation between Z_i and $X_i(s)$. In

particular, we set $\mathrm{Cov}(z_{ik}, v_{i11}) = 0.1$ for all $1 \leq k \leq 4$. We took 101 dense observations on grid points $s_{im} = 0.01m - 0.01$ with $1 \leq m \leq 101$ for each curve $X_i(s)$. The noisy observations were obtained by $W_{im} = X_i(s_{im}) + \epsilon_{im}$ for $1 \leq i \leq n$, where ϵ_{im} were independently generated from $N(0, 0.5)$.

We considered sample sizes of $n = 200, 500,$ and 1000 for each censoring rate and generated 100 datasets for each case. We fitted the FLCRM to each simulated dataset. To select the number of principal components, we use AIC method. We calculated the relative mean square errors of estimated coefficient function and parameter estimates according to $\mathrm{RMSE}_\beta = \int \{\widehat{\beta}(s) - \beta_0(s)\}^2 ds / \int \beta_0^2(s) ds$ and $\mathrm{RMSE}_\gamma = ||\widehat{\gamma} - \gamma_0||^2 / ||\gamma_0||^2$. These relative mean square error values characterize the accuracy of $\widehat{\beta}(\cdot)$ and $\widehat{\gamma}$. We also report the prediction performance by using the concordance index (Harrell et al., 1996; Heagerty and Zheng, 2005), which can be implemented using the R function "`concordance.index()`" in the R package "`survcomp`." To examine the effects of r_n on the estimation of parameters, we varied r_n from 1 to 10. We have also compared our method with the methods proposed in Gellar et al. (2015) and Qu et al. (2016). We include the estimation results for $n = 200$ and censoring rate 0.1 in Table 1. For all other eight settings, we include them in the supplementary document, Tables S1–S8.

The results reveal that the estimation of $\beta(s)$ is quite sensitive to r_n. Specifically, changing r_n can have a large effect on the estimation of $\beta(s)$. This indicates that selecting r_n by thresholding the percentage of variance explained would not work well in estimation since we do not know which $\mathrm{PV}(r_n)$ we should use. However, it turns out that the AIC tuning method performs reasonably well in selecting r_n, leading to good estimates of $\beta(s)$. In contrast, the estimation of γ and the predictive accuracy are quite robust to the selection of r_n. For the method of Gellar et al. (2015), it performs similarly as our method for all the scenarios. However, for the

method of Qu et al. (2016), it fails to estimate $\beta(s)$ and γ accurately. We guess that it may be caused by not using the traditional Newton–Raphson algorithm to obtain the estimator from penalized cox regression model. Specifically, they use the simplex search method of Lagarias et al. (1998), which is not guaranteed to converge to a local minimum, and therefore their method is not stable especially when the number of parameters is large.

3.2. Testing

We assess the Type I and II error rates of our score statistic by testing the hypotheses $H_0 : \beta_0(\cdot) = 0$ versus $H_1 : \beta_0(\cdot) \neq 0$. We used the same simulation method as that in Section 3.1 except that we set $\beta_0(s) = \beta_{C_1, C_2}(s) = 0.3[C_1\{\sin(\pi s) - \cos(\pi s) + \sin(3\pi s/10) - \cos(3\pi s) + \sin(5\pi s)/9 - \cos(5\pi s)/9 + \sin(7\pi s)/16 - \cos(7\pi s)/16 + \sin(9\pi s)/25 - \cos(9\pi s)/25\} + C_2(2\pi\sigma^2)^{-1/2} \exp\{-(2\sigma^2)^{-1}(s - 0.5)^2\}]$, where $C_1, C_2 \geq 0$ are scalars that control the degree of departure from H_0. We consider the censoring rates 10, 30, and 50% under sample sizes of $n = 200, 500,$ and 1000. We selected the number of principal components using percentage of variance explained and AIC method. In particular, we compare the type-I error performance by thresholding six different percentages of variance explained, that is, $\mathrm{PV}(r_n) = 0.70, 0.75, 0.80, 0.85, 0.90, 0.95$ and AIC method. For each simulated dataset, we calculated the score statistic and its associated p-value. The level of significance was set as 0.05. We used 5000 simulated datasets to estimate the Type I error rate when $C_1 = C_2 = 0$, that is, $\beta_{C_1, C_2}(\cdot) = 0$.

Table 2 summarizes type-I error rates of T_S at the nominal level of 5%. Inspecting Table 2 reveals that the Type I error rates are relatively accurate for all censoring rates and sample sizes when we use the percentages of variance explained. In contrast, when we use AIC selection method, the Type I error rates are significantly inflated. This indicates that we cannot use AIC method for testing.

Consequently, we only study the power analysis for thresholding percentages of variance explained method. In particular, we used 500 simulated datasets to estimate the power for two different alternative settings. The first alternative setting is to set $C_2 = 0$ and $C_1 = 0.1 \times j$ for $j = 1, \ldots, 10$. The second alternative setting is to set $C_1 = 0$ and $C_2 = 0.1 \times j$ for $j = 1, \ldots, 10$. We include simulation results when we use different thresholds $\mathrm{PV}(r_n) = 0.70, 0.75, 0.80, 0.85, 0.90,$ and 0.95 to select the number of FPCs. See Tables S9 and S10 of the supplementary document. From the results, we can see that the power of the test is quite robust to the choice of $\mathrm{PV}(r_n)$. Thus, we use $\mathrm{PV}(r_n) = 0.85$ throughout the article. Figure 1(a)–(c) give the statistical power of rejecting H_0 under the first set of alternative hypotheses and Figure 1(d)–(f) give the power of rejecting H_0 under the second setting. As expected, both increasing n and reducing the censoring rate improve the statistical power of rejecting the null hypothesis.

4. Alzheimers Disease Neuroimaging Initiative

Data used in the preparation of this article were obtained from the ADNI database (adni.loni.usc.edu). The detailed data description including hippocampus image data preprocessing and demographic information summary can be found in Section 2 in the supplementary document. The hippocampus

Table 1

Simulation results for the estimation and predictive accuracy of our estimation method when $n = 200$ and censoring rate is 0.1. We vary r_n from 1 to 10, and use AIC to select r_n, and we also compare with Gellar et al. (2015)'s method. The means of the estimates of $RMSE_\beta$, $RMSE_\gamma$, $\hat{r}_n$ and the concordance index with their standard errors in the parentheses were reported.

r_n	RMSE_β	RMSE_γ	Concordance index	$\hat{r}_n$
1	0.22(0.007)	0.24(0.018)	0.731(0.0004)	1
2	0.15(0.008)	0.25(0.019)	0.735(0.0005)	2
3	0.12(0.008)	0.26(0.02)	0.737(0.0005)	3
4	0.09(0.005)	0.26(0.019)	0.738(0.0003)	4
5	0.13(0.008)	0.26(0.019)	0.738(0.0003)	5
6	0.23(0.019)	0.27(0.02)	0.737(0.0004)	6
7	0.46(0.043)	0.28(0.02)	0.736(0.0005)	7
8	0.66(0.051)	0.29(0.02)	0.735(0.0005)	8
9	1.02(0.066)	0.29(0.02)	0.734(0.0005)	9
10	1.42(0.092)	0.29(0.02)	0.733(0.0005)	10
AIC	0.28(0.052)	0.26(0.019)	0.737(0.0005)	3.68(0.16)
Gellar	0.26(0.047)	0.26(0.019)	0.737(0.0004)	NA
Qu	3.70 (0.04)	4.27(0.07)	NA	NA

Table 2

Simulation results for T_S: the Type I error rates were reported at three censoring rates and three sample sizes. The number $0.70, 0.75, 0.80, 0.85, 0.90, 0.95$ in the first row denotes different $PV(r_n)$ values, and AIC denotes the AIC selection method. For each case, 5000 simulated datasets were used.

n	Censoring	0.7	0.75	0.80	0.85	0.90	0.95	AIC
200	0.1	0.06	0.059	0.0586	0.059	0.06	0.066	0.173
	0.3	0.0636	0.062	0.0626	0.0622	0.0664	0.0698	0.1782
	0.5	0.064	0.064	0.0644	0.0642	0.0672	0.07	0.1808
500	0.1	0.057	0.0514	0.0514	0.0514	0.0554	0.0574	0.162
	0.3	0.057	0.0566	0.0566	0.057	0.0542	0.0568	0.1578
	0.5	0.0574	0.0562	0.0562	0.0566	0.056	0.0598	0.1568
1000	0.1	0.055	0.0556	0.0556	0.0556	0.0554	0.058	0.1514
	0.3	0.0502	0.0556	0.0556	0.0556	0.056	0.0566	0.1504
	0.5	0.0536	0.0546	0.0546	0.0544	0.0574	0.0572	0.1548

is one of the key brain areas affected by AD. We consider the clinical and imaging measures of 373 MCI individuals in ADNI1 and use them to predict the time of conversion from MCI to AD, as well as to assess the predictive role of hippocampus surface morphology at baseline. Among the 373 MCI individuals, 161 MCI individuals progressed to AD before study completion and the remaining 212 MCI individuals did not convert to AD prior to study end. Thus, the time of conversion from MCI to AD can be treated as time-to-event data.

We fitted the FLCRM model to the ADNI dataset. The scalar covariate includes Gender ($1 = $ Male; $2 = $ Female),

Handedness ($1 = $ Right; $2 = $ Left), Marital Status ($1 = $ Married; $2 = $ Widowed; $3 = $ Divorced; $4 = $ Never married), Education length, Retirement ($1 = $ Yes; $0 = $ No), Age, the APOE genetic covariates with two SNPs, and the ADAS-Cog score. The two SNPs in APOE together define a 3-allele haplotype, namely, the $\epsilon 2$, $\epsilon 3$, and $\epsilon 4$ variants. For the categorical variables, we introduce dummy variables to represent them, resulting in a design matrix Z with dimension $(n, p) = (373, 12)$. For the functional predictors, we used hippocampal radial distances of 30,000 surface points on the left and right hippocampus surfaces. The radial distance is defined as the distance between the medial core of the
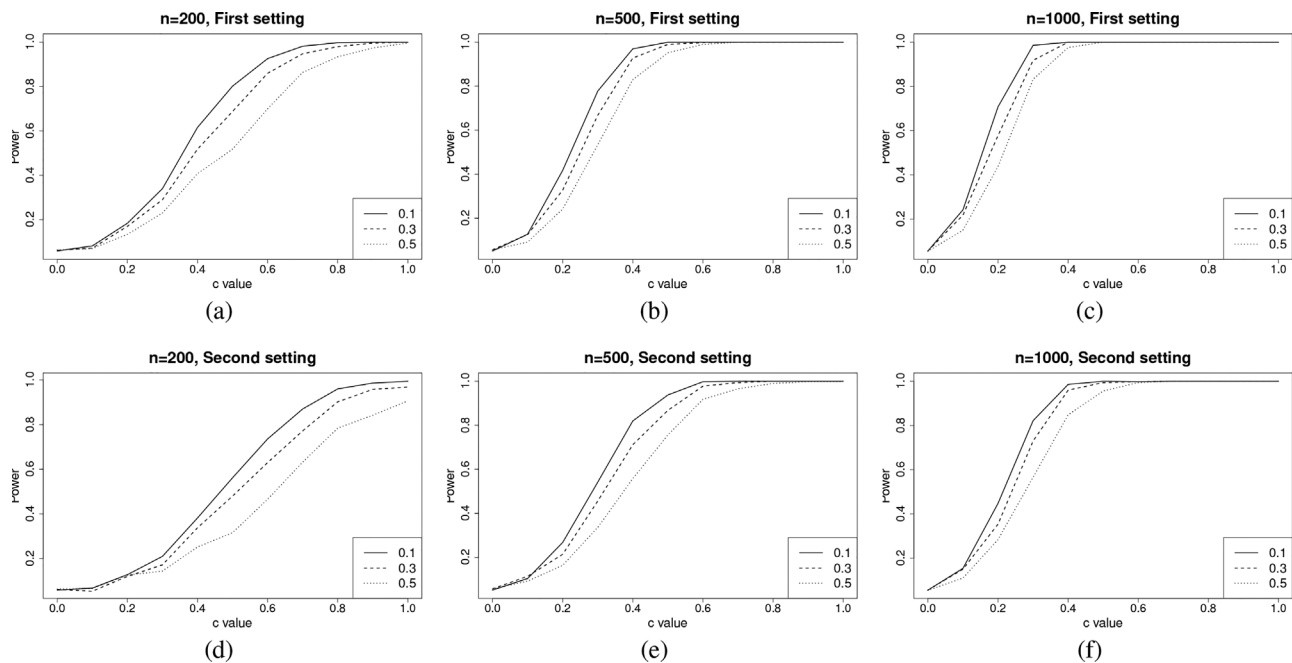

Figure 1. Simulation results for evaluating the rejection rate of T_S when $PV(r_n) = 0.85$: panels (a), (b), and (c), respectively, correspond to the changes in power for sample sizes of $n = 200$, 500, and 1000 under the first setting of alternative hypothesis; panels (d), (e), and (f), respectively, correspond to the changes in power for sample sizes of $n = 200$, 500, and 1000 under the second setting of alternative hypothesis. The solid, dashed, and dotted lines correspond to censoring rate of 0.1, 0.3, and 0.5, respectively.
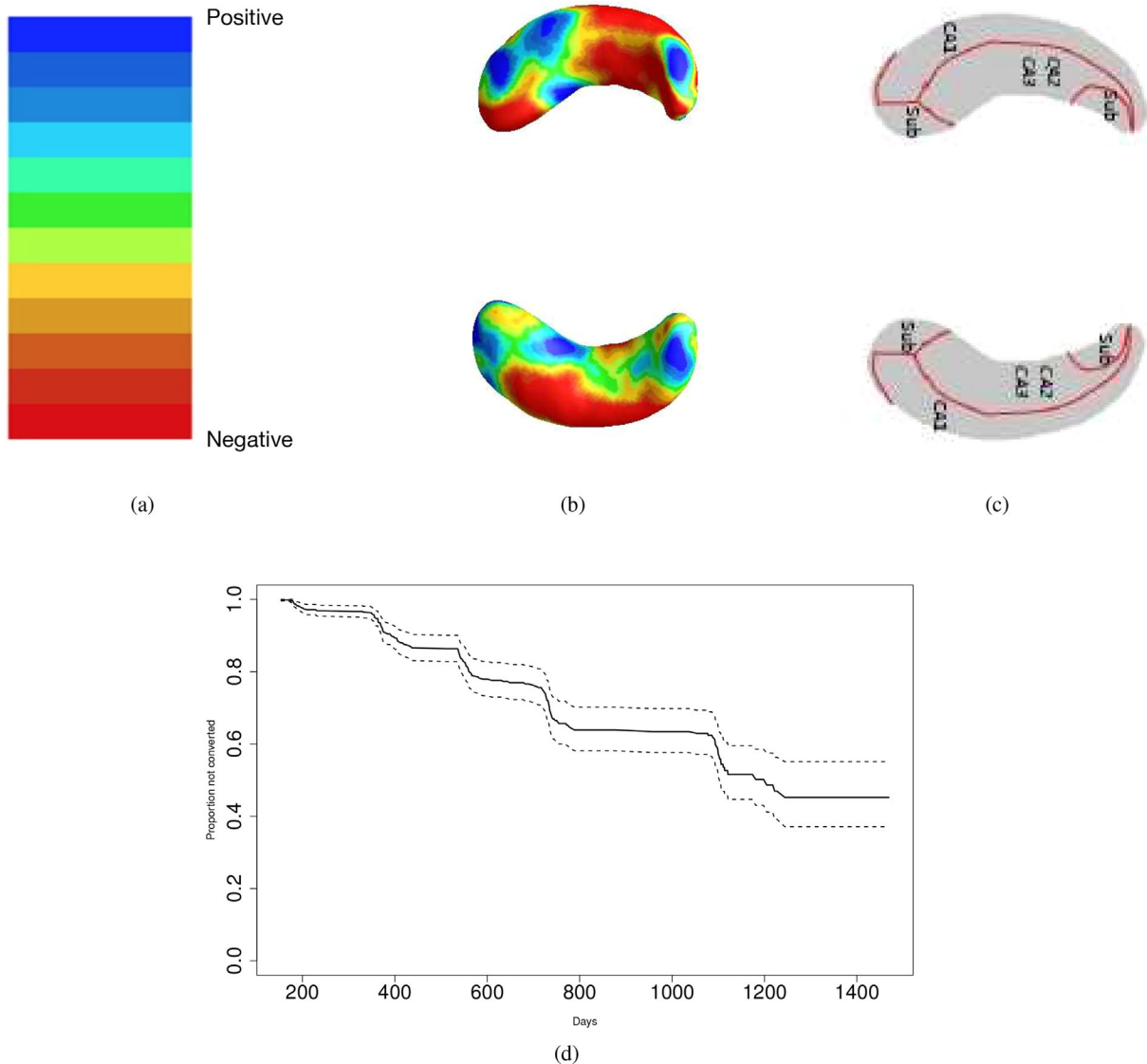
Figure 2. ADNI data analysis results: panel (a) is the color bar illustration, panel (b) contains the estimated coefficient functions $\beta(s)$, panel (c) is the hippocampal subfields, and panel (d) is the survival function for the time of MCI to AD progression. This figure appears in color in the electronic version of this article.

hippocampus and the corresponding vertex, and it is a summary statistic of the hippocampal shape and size. We applied FPCA to such surfaces to estimate the principal component scores. We selected the top 34 FPCs that explain 85% of the total variance. We used our score test statistic T_S to test the null hypothesis of $H_0 : \beta(\cdot) = 0$, and its associated p-value is 5×10^{-6}. This result may indicate a significant non-zero effect of the hippocampus radial distance surface data on the conversion time. We have performed the back-testing procedure to validate our testing result. In particular, we considered top 34 FPCs and applied our method to test whether each PC is significant compared with the null model with only clinical covariates. Then we adjust for multiple comparisons by using Bonferroni correction and compare all of the p-values with 0.05/34. The first principal component is still significant with its p-value 8×10^{-7}. Therefore, the

signal is very strong, validating our testing results. We also performed the sensitivity analysis of our testing procedure by using different percentages of variance explained to select r_n. In particular, we consider 70, 75, 80, 90, and 95%, and the corresponding p-values are 2×10^{-5}, 2×10^{-6}, 2×10^{-6}, 3×10^{-8}, and 7×10^{-7}, respectively. Therefore, all the p-values support a significant non-zero effect even after the use of Bonferroni adjustment for multiple comparisons.

We then used the estimation procedure of FLCRM to estimate γ and $\beta_0(\cdot)$. The number of FPCs is chosen to be $\hat{r}_n = 20$ by AIC. We fitted the FLCRM and applied the method in Grambsch and Therneau (1994) to test the proportional hazards assumption for a Cox regression model fit. This can be implemented by the "`cox.zph()`" function in the R package `survival`. The p-value is 0.177, which suggests that the proportional hazards assumption is not violated. Figure 2(b)

Table 3
ADNI data analysis: the estimates of γ and their standard errors

	γ_1	γ_2	γ_3	γ_4	γ_5	γ_6	γ_7	γ_8	γ_9	γ_{10}	γ_{11}	γ_{12}
Estimate	0.390	0.225	0.112	−0.676	0.132	−0.030	0.419	−0.035	0.407	0.204	−0.438	0.132
Standard error	0.197	0.324	0.298	0.409	0.786	0.030	0.226	0.015	0.407	0.458	0.202	0.020
p-values	0.048	0.488	0.707	0.099	0.866	0.320	0.064	0.017	0.317	0.656	0.030	1×10^{-10}

includes the estimated coefficient functions. We have plotted the hippocampal subfields in Figure 2(c). Inspecting Figure 2 reveals that the subfield of CA1 on both hippocampi has negative effects on the hazard function, indicating that the thicker these areas on the hippocampus are, the shorter the time is to covert to AD. Table 3 presents the estimated γ and their standard errors and p-values. We observe that ADAS-Cog score is significant, whereas the APOE genes are not after Bonferroni correction. This coincides with the results in Da et al. (2014), which show that a combination of spatial patterns of brain atrophy and ADAS-Cog offers good predictive power of conversion from MCI to AD, whereas APOE genotype does not significantly improve prediction. Our findings support prior MRI studies of volumetric hippocampal changes in prodromal AD (Dickerson and Wolk, 2013), and extend them by finding that the possible prognostic value of adding hippocampus surface data may be superior to that provided by routine clinical cognitive testing data.

We plotted the survival function with the 95% point-wise confidence interval in Figure 2(d), where the values of all the covariates are taken as the mean value of the covariates. In addition, we performed some sensitivity analysis of r_n. In particular, we consider $r_n = 17, 18, 19, 21, 22,$ and 23 and then we estimate γ and $\beta_0(\cdot)$. The estimated coefficient functions are presented in Figure S1 in the supplementary document. Compared with Figure 2(b), our estimated coefficient functions are quite robust to the choice of r_n.

We investigate the predictive performance of the FLCRM method. We also compared our model with a reduced model without the hippocampus surface data, and we fitted a Cox regression with only the scalar covariates. We randomly selected 200 subjects as the training data. We calculated the concordance index using the remaining 173 data points as the test data. We repeated this step for 100 times in order to obtain the mean of the concordance indices and its standard error for the two fitted models. The results show that the mean of the concordance indices is 0.68 (0.003) for our FLCRM, whereas the mean is 0.65 (0.003) for the reduced model, indicating the better prediction of our model. We also want to note that the concordance index with 0.68 is not very high, and there is still a lot of unmeasured variability driving progression to AD. We also compared with a new model by replacing the surface data by the hippocampal size, and the mean of the concordance indices is 0.69 (0.003). Although the summary statistics provide similar prediction power, the use of the whole hippocampal surface allows us to localize the effects of hippocampal subregions on conversion from MCI to AD.

5. Supplementary Material

Web Appendices, Tables, and Figures referenced in Sections 2–4 and the computer code are available with this article at the *Biometrics* website on Wiley Online Library.

References

Cai, T., Tonini, G., and Lin, X. (2011). Kernel machine approach to testing the significance of multiple genetic markers for risk prediction. *Biometrics* **67**, 975–986.

Cardot, H., Ferraty, F., Mas, A., and Sarda, P. (2003). Testing hypotheses in the functional linear model. *Scandinavian Journal of Statistics* **30**, 241–255.

Cox, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society, Series B* **34**, 187–220.

Cox, D. R. (1975). Partial likelihood. *Biometrika* **62**, 269–276.

Da, X., Toledo, J., Zee, J., Wolk, D. A., Xie, S. X., Ou, Y., et al. (2014). Integration and relative value of biomarkers for prediction of mci to ad progression: Spatial patterns of brain atrophy, cognitive scores, apoe genotype, and csf markers. *Neuroimage: Clinical* **4**, 164–173.

Dickerson, B. C. and Wolk, D. (2013). Biomarker-based prediction of progression in mci: Comparison of ad-signature and hippocampal volume with spinal fluid amyloid-β and tau. *Frontiers in Aging Neuroscience* **5**, 55.

Efron, B. (1977). The efficiency of Cox's likelihood function for censored data. *Journal of the American Statistical Association* **72**, 557–565.

Fan, J. and Gijbels, I. (1996). *Local Polynomial Modelling and Its Applications.* London: Chapman and Hall.

Gellar, J. E., Colantuoni, E., Needham, D. M., and Crainiceanu, C. M. (2015). Cox regression models with functional covariates for survival data. *Statistical Modelling* **15**, 256–278.

Grambsch, P. M. and Therneau, T. M. (1994). Proportional hazards tests and diagnostics based on weighted residuals. *Biometrika* **81**, 515–526.

Hall, P. and Horowitz, J. L. (2007). Methodology and convergence rates for functional linear regression. *The Annals of Statistics* **35**, 70–91.

Hall, P. and Hosseini-Nasab, M. (2006). On properties of functional principal components analysis. *Journal of the Royal Statistical Society, Series B* **68**, 109–126.

Harrell, F. E., Lee, K. L., and Mark, D. B. (1996). Tutorial in biostatistics multivariable prognostic models: Issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Statistics in Medicine* **15**, 361–387.

Heagerty, P. J. and Zheng, Y. (2005). Survival model predictive accuracy and ROC curves. *Biometrics* **61**, 92–105.

Huang, J., Sun, T., Ying, Z., Yu, Y., and Zhang, C. (2013). Oracle inequalities for the lasso in the cox model. *The Annals of Statistics* **41**, 1142–1165.

Ibrahim, J. G., Chen, M.-H., and Sinha, D. (2001). *Bayesian Survival Analysis*. New York: Springer-Verlag Inc.

Johnstone, I. M. and Lu, A. Y. (2009). On consistency and sparsity for principal components analysis in high dimensions. *Journal of the American Statistical Association* **104**, 682–693.

Kalbfleisch, J. D. and Prentice, R. L. (2002). *The Statistical Analysis of Failure Time Data*. 2nd edition. New York: Wiley.

Kong, D., Staicu, A.-M., and Maity, A. (2016). Classical testing in functional linear models. *Journal of Nonparametric Statistics* **28**, 813–838.

Lagarias, J. C., Reeds, J. A., Wright, M. H., and Wright, P. E. (1998). Convergence properties of the nelder–mead simplex method in low dimensions. *SIAM Journal on Optimization* **9**, 112–147.

Lee, E., Zhu, H., Kong, D., Wang, Y., Giovanello, K. S., and Ibrahim, J. G. (2015). BFLCRM: A bayesian functional linear cox regression model for predicting time to conversion to alzheimer's disease. *The Annals of Applied Statistics* **9**, 2153–2178.

Li, J. and Ma, S. (2013). *Survival Analysis in Medicine and Genetics*. Chapman & Hall/CRC Biostatistic Series, Boca Raton, FL: CRC Press, Taylor & Francis Group, LLC.

Li, S., Okonkwo, O., Albert, M., and Wang, M.-C. (2013). Variation in variables that predict progression from mci to ad dementia over duration of follow-up. *American Journal of Alzheimer's Disease* **1**, 12–28.

Li, Y., Wang, N., and Carroll, R. J. (2010). Generalized functional linear models with semiparametric single-index interactions. *Journal of the American Statistical Association* **105**, 621–633.

Morris, J. S. (2015). Functional regression. *Annual Reviews of Statistics and Its Application* **2**, 321–359.

Petersen, R. C. (2004). Mild cognitive impairment as a diagnostic entity. *Journal of Internal Medicine* **256**, 183–194.

Qu, S., Wang, J., and Wang, X. (2016). Optimal estimation for the functional cox model. *Annals of Statistics* **44**, 1708–1738.

Reiss, P. T. and Ogden, R. T. (2007). Functional principal component regression and functional partial least squares. *Journal of the American Statistical Association* **102**, 984–996.

Reiss, P. T. and Ogden, R. T. (2010). Functional generalized linear models with images as predictors. *Biometrics* **66**, 61–69.

Risacher, S. L., Saykin, A. J., Wes, J. D., Shen, L., Firpi, H. A., and McDonald, B. C. (2009). Baseline MRI predictors of conversion from MCI to probable AD in the ADNI cohort. *Current Alzheimer Research* **6**, 347–361.

Xu, J. (2012). High-dimensional cox regression analysis in genetic studies with censored survival outcomes. *Journal of Probability and Statistics* **2012**, 1–14.

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

Zhang, J. and Chen, J. (2007). Statistical inference for functional data. *The Annals of Statistics* **35**, 1052–1079.