Penalized estimation for varying coefficient additive hazards models

Journal Title
XX(X):1–16

DOI: 10.1177/ToBeAssigned
www.sagepub.com/
SAGE

Abstract

Varying coefficient models are commonly used to capture intricate interaction effects among covariates in regression models, allowing for the modification of one covariate's effect by another. Although these models offer increased flexibility, they also introduce greater estimation and computational complexity as a trade-off. This complexity is particularly evident in genomic studies, where the covariates are often high-dimensional, rendering conventional estimation methods inapplicable. In this paper, we study a penalized estimation method for the varying coefficient additive hazards model. We adopt the group lasso penalty along with the kernel smoothing technique to estimate the varying coefficients. In contrast to existing kernel methods, which only use a "local" neighborhood of subjects to estimate the varying coefficient function at any given point, the proposed method takes a "global" approach that incorporates all subjects and is more efficient. Through extensive simulation studies, we demonstrate that the proposed method produces interpretable results with satisfactory predictive performance. We provide an application to a major cancer genomic study.

Keywords

Censored data, kernel smoothing, semiparametric model, survival analysis, variable selection

1 Introduction

In biomedical research, key issues include accurate prediction of disease outcomes and effective assignment of treatment strategies based on individual patient characteristics. To address these issues, it is crucial to construct models that capture the intricate associations between survival outcomes, such as time to death or disease progression, and different risk factors. In cancer studies, survival models are widely employed to study the effects of genomic features on cancer progression. Among various survival models, the proportional hazards model [1] and the additive hazards model [2,3] are popular choices for assessing the association between genomic features and survival times. The proportional hazards model [1] assumes that the covariates act multiplicatively on the hazard function. Its popularity is mainly due to the simple interpretation of the regression coefficients as hazards ratios without assuming a specific event time distribution. As an alternative to the proportional hazards model, the additive hazards model [3] assumes that the hazard function of an event time, given a time-independent covariate vector $\boldsymbol{X}$, takes

the form

$$\lambda(t \mid \boldsymbol{X}) = \lambda(t) + \boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{X}, \tag{1}$$

where $\lambda(\cdot)$ is an unspecified positive function, and $\boldsymbol{\beta}$ is a vector of regression coefficients. In this model, the hazard function is a linear combination of the covariate vector, with the covariates acting additively on an unknown baseline hazard function. Since the two models assume different structures of association between the covariates and event time, both are of practical interest, especially in the data exploration stage. While the choice between the models should primarily be based on the scientific question of interest and the empirical fit of data, one advantage of the additive hazards model is the availability of closed-form estimators of the regression parameters.

In certain scenarios, the effect of risk factors on cancer progression can differ among patient subgroups, such as different age groups. For instance, in renal cell carcinoma, most clinical factors demonstrate stronger prognostic value in younger patients. Understanding how each risk factor influences cancer progression across different age groups is essential for a more precise prognosis.[4] There has been a growing interest in utilizing aging-related gene signatures for prognosis in various cancer types.[5–8] In addition, the effect of certain copy number variations on survival may vary among patients exposed to different drugs.[9] Assessing the effect of each risk factor across subgroups is crucial for developing personalized treatment strategies. One approach to capture distinct covariate effects is by representing them as functions of a relevant variable associated with the risk factors. In such instances, a varying coefficient additive hazards model can be considered. We assume that the hazard function of an event time, given the covariates W and $\boldsymbol{X}$, takes the form

$$\lambda(t \mid W, \boldsymbol{X}) = \lambda(t) + \boldsymbol{\beta}(W)^{\mathrm{T}} \boldsymbol{X}, \tag{2}$$

where $\boldsymbol{\beta}(w)$ is a vector of coefficient functions that represents the effect of $\boldsymbol{X}$ at $W = w$. Model (2) allows for a dynamic representation of the association between risk factors and the event time, specifically accounting for the variations observed across different subgroups indexed by W.

Lin and Ying[3] proposed an estimating function for $\boldsymbol{\beta}$ under model (1), which mimics the martingale feature of the partial likelihood score function for $\boldsymbol{\beta}$ under the proportional hazards model. A consistent kernel estimator for $\boldsymbol{\beta}(w)$ can be obtained by incorporating a kernel weight for each observation to the Lin and Ying estimator. Since only a subset of subjects near the neighborhood of w is effectively used for the estimation of $\boldsymbol{\beta}(w)$, we refer to this method as a "local" kernel estimation method. This local kernel estimator can be inefficient because it implicitly assumes that the baseline hazard function varies with W, and only a subset of subjects is used to profile out the baseline hazard function. Ng and Wong[10] proposed a global kernel estimation method to mitigate the efficiency loss. In particular, the global method estimates the varying coefficient function over a range of covariate values simultaneously and thus uses all subjects to profile out the baseline hazard function. It has been demonstrated that the global kernel estimator exhibits improved efficiency compared to existing local kernel estimators.

In some applications, such as genomic studies, $\boldsymbol{X}$ may be high-dimensional, or the primary interest may lie in selecting a subset of important risk factors. The global estimator of Ng and Wong[10] quickly becomes infeasible as the number of covariates grows; the dimensionality issue is exacerbated by the fact that each covariate is associated with a number of regression parameters. Therefore, it is crucial to explore variable selection methods based on the global estimator for varying coefficient additive hazards models.

For a general additive hazards model, several penalization methods have been proposed for simultaneous variable selection and estimation.[11–15] Qu et al.[16] extended these penalization methods to accommodate varying coefficients in additive hazards models, by combining the kernel smoothing technique and the weighted lasso penalty to perform variable selection and identify local sparse effects. However, it is worth noting that the kernel estimator proposed by Qu et al.[16] is constructed locally, assuming that all parameters, including the baseline hazard function, are functions of the variable W.

In this paper, we extend the global kernel estimation methods of Ng and Wong[10] by incorporating penalization on the varying covariate effects. This allows for the identification of covariates (with potential interactions with W) that impact survival outcomes. The penalized estimation of the varying coefficient additive hazards model presents two main challenges. First, the inclusion of a penalty imposes additional computational challenges and renders a closed-form solution unavailable. Second, obtaining valid statistical inference for the varying coefficient functions could be challenging and the introduction of penalties further complicates the inference procedures. Because the penalized estimator is not asymptotically normal, standard statistical inference procedures cannot be applied.

The remainder of this paper is organized as follows. Section 2 describes the proposed penalization method. Section 3 demonstrates the empirical performance of the proposed method through simulation studies. Section 4 illustrates an application of the proposed method to cancer genomic data. Section 5 includes some concluding remarks and possible directions for extensions.

2 Estimation

2.1 Preliminaries

We introduce relevant notations and provide a brief description of the existing estimation methods for the (varying coefficient) additive hazards model. We focus on an event time $\widetilde{T}$, which follows the hazard function specified in model (2). This event time is potentially right-censored, and we denote the censoring time as C. We assume that C and $\widetilde{T}$ are conditionally independent given $(W, \boldsymbol{X})$. Let $T \equiv \min(\widetilde{T}, C)$ be the observed time and $\Delta \equiv I(\widetilde{T} \leq C)$ be the event indicator. For a sample of size n, we observe $(T_i, \Delta_i, W_i, \boldsymbol{X}_i)$ for $i = 1, \ldots, n$, where $\boldsymbol{X}_i = (X_{i1}, \ldots, X_{ip})$. Also, we define $N_i(t) = \Delta_i I(T_i \leq t)$ as the observed event process and $Y_i(t) = I(T_i \geq t)$ as the at-risk process for the i-th subject at time t.

When $\boldsymbol{\beta}$ is constant as in model (1), we can estimate $\boldsymbol{\beta}$ using the Lin and Ying estimator:

$$\left[\sum_{i=1}^{n} \int_0^\tau Y_i(t)\{\boldsymbol{X}_i - \overline{\boldsymbol{X}}(t)\}^{\otimes 2}\, \mathrm{d}t\right]^{-1} \sum_{i=1}^{n} \int_0^\tau \{\boldsymbol{X}_i - \overline{\boldsymbol{X}}(t)\}\, \mathrm{d}N_i(t),$$

where τ is the maximum follow-up time, $\overline{\boldsymbol{X}}(t) \equiv \sum_{i=1}^n \boldsymbol{X}_i Y_i(t) / \sum_{i=1}^n Y_i(t)$ is the average covariate vector among subjects at risk at time t with the convention that $0/0 = 0$, and for any vector $\boldsymbol{a}$, $\boldsymbol{a}^{\otimes 2} = \boldsymbol{a}\boldsymbol{a}^{\mathrm{T}}$.

When $\boldsymbol{\beta}$ depends on W as in model (2), an intuitive way to estimate $\boldsymbol{\beta}(\cdot)$ is to include the kernel function as a weight to each observation in the Lin and Ying estimator, resulting in an estimator of $\boldsymbol{\beta}(w)$

given by

$$\left[\sum_{i=1}^{n}\int_0^\tau K_h(W_i-w)Y_i(t)\left\{\boldsymbol{X}_i-\frac{\sum_{i=1}^n K_h(W_i-w)Y_i(t)\boldsymbol{X}_i}{\sum_{i=1}^n K_h(W_i-w)Y_i(t)}\right\}^{\otimes 2}\mathrm{d}t\right]^{-1}$$
$$\times\sum_{i=1}^{n}\int_0^\tau K_h(W_i-w)\left\{\boldsymbol{X}_i-\frac{\sum_{i=1}^n K_h(W_i-w)Y_i(t)\boldsymbol{X}_i}{\sum_{i=1}^n K_h(W_i-w)Y_i(t)}\right\}\mathrm{d}N_i(t),$$

where $K_h(x)\equiv K(x/h)/h$ is a kernel function with bandwidth h. This kernel estimator solves a set of estimating equations that includes subjects in the neighborhood of w. This estimator, based on a small local neighborhood of subjects, can be viewed as a "local" kernel estimator. When $K_h(W_i-w)$ is constant for all w, the local kernel estimator reduces to the Lin and Ying estimator. While the local estimator is natural and easy to compute, it could suffer efficiency loss because only a small neighborhood of subjects are used to simultaneously estimate the varying coefficient function at any given value and profile out the baseline hazard function. It is crucial to consider the shared nature of the baseline hazard function among all subjects in order to enhance the efficiency of parameter estimation.

To take into account the common baseline hazard function across W, Ng and Wong[10] proposed a global kernel estimation method. They proposed to approximate the coefficient function $\boldsymbol{\beta}(\cdot)$ on a grid $(w_1,\ldots,w_m)$ and estimate the function values $(\boldsymbol{\beta}(w_1)^{\mathrm{T}},\ldots,\boldsymbol{\beta}(w_m)^{\mathrm{T}})^{\mathrm{T}}$ by $\widehat{\boldsymbol{\beta}}\equiv\boldsymbol{V}^{-1}\boldsymbol{b}$ (which is equivalent to minimizing $\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{V}\boldsymbol{\beta}/2-\boldsymbol{b}^{\mathrm{T}}\boldsymbol{\beta}$), where $\boldsymbol{V}$ is a block matrix with $\boldsymbol{V}(w_k,w_\ell)$ as its (k,ℓ)th block, $\boldsymbol{b}=(\boldsymbol{b}(w_1)^{\mathrm{T}},\ldots,\boldsymbol{b}(w_m)^{\mathrm{T}})^{\mathrm{T}}$,

$$\begin{aligned}
\boldsymbol{b}(w_k)=&\frac{1}{nm}\sum_{i=1}^{n}\int_0^\tau\left\{K_h(W_i-w_k)\boldsymbol{X}_i-\sum_{j=1}^{m}K_h(W_i-w_j)\overline{\boldsymbol{X}}(t,w_k)\right\}\mathrm{d}N_i(t),\\
\boldsymbol{V}(w_k,w_\ell)=&\frac{I(w_k=w_\ell)}{nm}\sum_{i=1}^{n}\int_0^\tau K_h(W_i-w_k)Y_i(t)\boldsymbol{X}_i^{\otimes 2}\,\mathrm{d}t\\
&-\frac{1}{nm}\sum_{i=1}^{n}\int_0^\tau\sum_{j=1}^{m}K_h(W_i-w_j)Y_i(t)\overline{\boldsymbol{X}}(t,w_k)\overline{\boldsymbol{X}}(t,w_\ell)^{\mathrm{T}}\,\mathrm{d}t,\\
\overline{\boldsymbol{X}}(t,w_k)=&\frac{\sum_{i=1}^n K_h(W_i-w_k)Y_i(t)\boldsymbol{X}_i}{\sum_{i=1}^n\sum_{j=1}^m K_h(W_i-w_j)Y_i(t)}.
\end{aligned}$$

This estimator is motivated by first considering the Lin and Ying estimator of model (2) under a discrete W and then suitably incorporating kernel weights to the estimator. For the points off the grid, linear interpolation can be adopted to approximate the corresponding function values. When W is a categorical variable, we can simply replace the kernel function $K_h(W-w)$ by the indicator function $I(W=w)$, or equivalently setting the bandwidth h to 0.

2.2 Penalized estimation

To handle a high-dimensional $\boldsymbol{X}$ or to promote sparsity, we may impose a lasso penalty [17] to estimate $\boldsymbol{\beta}$. This penalized estimation involves minimizing the following penalized loss function:

$$\frac{1}{2}\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{V}\boldsymbol{\beta} - \boldsymbol{b}^{\mathrm{T}}\boldsymbol{\beta} + \lambda \sum_{j=1}^{p}\sum_{k=1}^{m} \left|\beta_j(w_k)\right|,$$

where $\lambda \geq 0$ is a tuning parameter. This approach allows for the identification of local sparse effects. Although it facilitates variable selection, the noise variable is more likely to be selected as the covariate effects corresponding to the same covariate (at different grid points) are penalized separately. To yield a more interpretable model, we apply a group lasso penalty [18] and estimate $\boldsymbol{\beta}$ by minimizing the following penalized loss function:

$$\frac{1}{2}\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{V}\boldsymbol{\beta} - \boldsymbol{b}^{\mathrm{T}}\boldsymbol{\beta} + \lambda \sum_{j=1}^{p} \|\boldsymbol{\beta}_j\|_2, \tag{3}$$

where $\boldsymbol{\beta}_j = (\beta_j(w_1), \ldots, \beta_j(w_m))^{\mathrm{T}}$. The group lasso penalty term promotes sparsity in a group level, such that the coefficients corresponding to the same covariate, namely $\beta_j(w_1), \ldots, \beta_j(w_m)$, will either be shrunk to all zero or all nonzero.

The group coordinate descent method [19] can be adopted to compute the estimator of $\boldsymbol{\beta}$. Since the penalized loss function is the sum of convex functions and is separable, the group coordinate descent can be used for its optimization. The group coordinate descent algorithm optimizes the penalized loss function by iteratively updating a group of covariate effects at a time and cycling through all the groups until convergence. In particular, for the j-th group of covariate effects, $\boldsymbol{\beta}_j$, the subgradient equations of (3) are

$$\sum_{\ell=1}^{p} V_{\ell j}(w_k, w_k)\beta_\ell(w_k) + \sum_{\ell \neq k}\sum_{i=1}^{p} V_{ij}(w_\ell, w_k)\beta_i(w_\ell) - b_j(w_k) + \lambda s_k = 0 \tag{4}$$

for $k = 1, \ldots, m$, where $s_k = \beta_j(w_k)/\|\boldsymbol{\beta}_j\|_2$ for $\boldsymbol{\beta}_j \neq \mathbf{0}$ and $(s_1, \ldots, s_m)$ satisfies $\sum_{k=1}^{m} s_k^2 \leq 1$ for $\boldsymbol{\beta}_j = \mathbf{0}$. If

$$c_j \equiv \frac{1}{\lambda^2}\sum_{k=1}^{m}\left\{ b_j(w_k) - \sum_{\ell=1}^{m}\sum_{i \neq j} V_{ij}(w_\ell, w_k)\beta_i(w_\ell) \right\}^2 = \sum_{k=1}^{m} s_k^2 \leq 1,$$

then $\boldsymbol{\beta}_j = \mathbf{0}$ is the solution to (4). Otherwise, we solve for $\boldsymbol{\beta}_j$ in (4) with other parameters fixed at their current estimates.

For fixed λ, the computation algorithm is outlined as follows:

Step 1. Initialize $\widehat{\boldsymbol{\beta}}$.

Step 2. For $j = 1, \ldots, p$,

Step 2(a). Evaluate c_j;

Step 2(b). If $c_j \leq 1$, then set $\widehat{\boldsymbol{\beta}}_j = \mathbf{0}$. If $c_j > 1$, then set $\widehat{\boldsymbol{\beta}}_j = (\boldsymbol{V}_{jj} + \lambda \boldsymbol{I})^{-1}\big(\boldsymbol{b}_j - \sum_{k \neq j} \boldsymbol{V}_{jk}\boldsymbol{\beta}_k\big)$.

Step 3. Repeat Step 2 until convergence.

We propose to select the tuning parameter λ using cross-validation, with the cross-validation loss set to be the unpenalized version of the loss function (3). To ensure that the magnitude of penalty is comparable across various covariate effects, we suggest to standardize each component of $\boldsymbol{X}$ before performing the penalized estimation.

2.3 Decorrelated score function and inference

The (group) lasso estimator, while effective at inducing sparsity, exhibits asymptotic bias due to excessive shrinkage to the nonzero regression coefficients. For more accurate estimation, we perform a debiasing step based on the decorrelated score function.[20] This approach involves constructing a decorrelated score function that is uncorrelated with the nuisance score function (the derivative of the loss function with respect to the parameters not directly of interest). The resulting debiased estimator, derived from solving the decorrelated score function, is asymptotically normal and consistent. The study of a debiased estimator is particularly important because it facilitates the development of a statistical inference procedure. Since the penalized estimator is not asymptotically normally distributed, it is challenging to perform statistical inference using the penalized estimator. By adjusting the penalized estimator using the decorrelated score function, statistical inference can be performed by inverting the asymptotic distribution of the debiased estimator.

Suppose that we are interested in making inference on $\beta_j(\cdot)$. We define $\boldsymbol{\theta} = \boldsymbol{\beta}_j = (\beta_j(w_1), \ldots, \beta_j(w_m))^{\mathrm{T}}$ as an m vector of parameters of interest and $\boldsymbol{\gamma} = (\boldsymbol{\beta}_1^{\mathrm{T}}, \ldots, \boldsymbol{\beta}_{j-1}^{\mathrm{T}}, \boldsymbol{\beta}_{j+1}^{\mathrm{T}}, \ldots, \boldsymbol{\beta}_p^{\mathrm{T}})^{\mathrm{T}}$ as an $m(p-1)$ vector of nuisance parameters. We also define the corresponding penalized estimators as $\widehat{\boldsymbol{\theta}}$ and $\widehat{\boldsymbol{\gamma}}$, respectively. Let $\boldsymbol{I}_{\theta\gamma}$ and $\boldsymbol{I}_{\gamma\gamma}$ be the corresponding partitions of $\mathrm{E}\{\nabla^2 \ell(\boldsymbol{\beta})\}$ with $\ell(\boldsymbol{\beta}) = \boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{V}\boldsymbol{\beta}/2 - \boldsymbol{b}^{\mathrm{T}}\boldsymbol{\beta}$. The decorrelated score function for $\boldsymbol{\theta}$ is $\nabla_\theta \ell(\boldsymbol{\theta}, \boldsymbol{\gamma}) - \boldsymbol{H}^{\mathrm{T}} \nabla_\gamma \ell(\boldsymbol{\theta}, \boldsymbol{\gamma})$ with $\boldsymbol{H}^{\mathrm{T}} = \boldsymbol{I}_{\theta\gamma}\boldsymbol{I}_{\gamma\gamma}^{-1}$. Equivalently, the decorrelated score function for $\boldsymbol{\theta}$ can be expressed as

$$\boldsymbol{S}(\boldsymbol{\theta}, \boldsymbol{\gamma}) = \begin{pmatrix} \boldsymbol{I} & -\boldsymbol{H}^{\mathrm{T}} \end{pmatrix} \left\{ \begin{pmatrix} \boldsymbol{V}_{\theta\theta} & \boldsymbol{V}_{\theta\gamma} \\ \boldsymbol{V}_{\gamma\theta} & \boldsymbol{V}_{\gamma\gamma} \end{pmatrix} \begin{pmatrix} \boldsymbol{\theta} \\ \boldsymbol{\gamma} \end{pmatrix} - \begin{pmatrix} \boldsymbol{b}_\theta \\ \boldsymbol{b}_\gamma \end{pmatrix} \right\},$$

where $\boldsymbol{V}_{\theta\theta}$, $\boldsymbol{V}_{\theta\gamma}$, $\boldsymbol{V}_{\gamma\theta}$, and $\boldsymbol{V}_{\gamma\gamma}$ represent the corresponding partitions of the matrix $\boldsymbol{V}$, and $\boldsymbol{b}_\theta$ and $\boldsymbol{b}_\gamma$ represent the corresponding partitions of the vector $\boldsymbol{b}$. The decorrelated score function for $\boldsymbol{\theta}$, $\widehat{\boldsymbol{S}}(\boldsymbol{\theta}, \boldsymbol{\gamma})$, can be estimated by plugging in $\widehat{\boldsymbol{H}} = \boldsymbol{V}_{\gamma\gamma}^{-1}\boldsymbol{V}_{\gamma\theta}$ into $\boldsymbol{S}(\boldsymbol{\theta}, \boldsymbol{\gamma})$. Let $\widehat{\boldsymbol{I}}_{\theta|\gamma} = \boldsymbol{V}_{\theta\theta} - \widehat{\boldsymbol{H}}^{\mathrm{T}}\boldsymbol{V}_{\gamma\theta}$ be the estimator of the partial information matrix. We define a debiased estimator $\widetilde{\boldsymbol{\theta}} = \widehat{\boldsymbol{\theta}} - \widehat{\boldsymbol{I}}_{\theta|\gamma}^{-1}\widehat{\boldsymbol{S}}(\widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\gamma}})$. This debiased estimator is obtained by deriving the first-order approximation of $\widehat{\boldsymbol{S}}(\boldsymbol{\theta}, \widehat{\boldsymbol{\gamma}})$ and solving this approximation equals zero. It can be verified that $\sqrt{nh}(\widetilde{\boldsymbol{\theta}} - \boldsymbol{\theta})$ is approximately $\mathrm{N}(\mathbf{0}, \boldsymbol{I}_{\theta|\gamma}^{-1})$.[20]

The debiased estimator, which is approximately normally distributed, can be used to construct a simultaneous confidence band for the varying coefficient function $\beta_j(\cdot)$ over the interval $[a, b]$ using

the perturbation method.[21] We derive a large-sample approximation to the distribution of

$$\mathcal{S}_j = \sup_{w\in[a,b]} \widehat{v}_j(w)\big|\widetilde{\beta}_j(w) - \beta_j(w)\big|,$$

where $\widetilde{\beta}_j(w)$ is the debiased estimator of $\beta_j(w)$ and $\widehat{v}_j(w)$ is a positive weight function that converges uniformly to a deterministic function. Note that $\sqrt{nh}(\widetilde{\boldsymbol{\theta}} - \boldsymbol{\theta}) = \sqrt{nh}(\widetilde{\beta}_j(w_1) - \beta_j(w_1), \ldots, \widetilde{\beta}_j(w_m) - \beta_j(w_m))^{\mathrm{T}}$ can be approximated by

$$\begin{aligned}
&\sqrt{nh}\widehat{\boldsymbol{I}}_{\theta|\gamma}^{-1}(\boldsymbol{I}, -\widehat{\boldsymbol{H}}^{\mathrm{T}})\boldsymbol{M} \\
\equiv&\sqrt{nh}\widehat{\boldsymbol{I}}_{\theta|\gamma}^{-1}(\boldsymbol{I}, -\widehat{\boldsymbol{H}}^{\mathrm{T}}) \begin{pmatrix} \begin{pmatrix} M_{nj}(w_1) \\ \vdots \\ M_{nj}(w_m) \end{pmatrix} \\ \begin{pmatrix} (M_{n1}(w_1), \ldots, M_{n(j-1)}(w_1), M_{n(j+1)}(w_1), \ldots, M_{np}(w_1))^{\mathrm{T}} \\ \vdots \\ (M_{n1}(w_m), \ldots, M_{n(j-1)}(w_m), M_{n(j+1)}(w_m), \ldots, M_{np}(w_m))^{\mathrm{T}} \end{pmatrix} \end{pmatrix},
\end{aligned}$$

where $M_{nj}(w) = n^{-1}\sum_{i=1}^{n}\int_0^\tau \{K_h(W_i - w)X_{ij} - \sum_{\ell=1}^{m} K_h(W_i - w_\ell)\overline{X}_j(t,w)\}\,\mathrm{d}M_i(t)$. Define $\widetilde{\boldsymbol{M}}$ as a stochastic perturbation of $\boldsymbol{M}$ by replacing $\mathrm{d}M_i(t)$ in $\boldsymbol{M}$ with $\mathrm{d}N_i(t)\psi_i$, where $\{\psi_i, i = 1, \ldots, n\}$ is an iid sample from the standard normal distribution. The distribution of $\mathcal{S}_j$ can be approximated by

$$\widetilde{\mathcal{S}}_j = \|\widehat{\boldsymbol{V}}_j\widehat{\boldsymbol{I}}_{\theta|\gamma}^{-1}(\boldsymbol{I}, -\widehat{\boldsymbol{H}}^{\mathrm{T}})\widetilde{\boldsymbol{M}}\|_\infty,$$

where $\widehat{\boldsymbol{V}}_j = \mathrm{diag}(\widehat{v}_j(w_1), \ldots, \widehat{v}_j(w_m))$. By repeatedly generating samples of $\{\psi_i, i = 1, \ldots, n\}$, we obtain an empirical distribution of $\widetilde{\mathcal{S}}_j$, which serves as an approximation of the distribution of $\mathcal{S}_j$. Let $c_{\alpha j}$ be the $(1-\alpha)$-th empirical quantile of $\widetilde{\mathcal{S}}_j$, where $0 < \alpha < 1$. A $(1-\alpha)$ confidence band for $\{\beta_j(w), w \in [a,b]\}$ can be constructed as

$$\{\widetilde{\beta}_j(w) \pm c_{\alpha j}\widehat{v}_j(w)^{-1}, a \le w \le b\}.$$

For $\widehat{v}_j(w_k)$, we set $\{\widehat{v}_j(w_k)\}^{-1}$ to be the estimated standard error of $\widetilde{\beta}_j(w_k)$, which is the k-th diagonal element of $\widehat{\boldsymbol{I}}_{\theta|\gamma}^{-1/2}$.

In the above procedure, $\boldsymbol{H}$ is estimated without regularization, and the estimator is valid when mp is sufficiently smaller than n. In the case that mp is moderately large, we can estimate $\boldsymbol{H}$ by a Dantzig-type estimator.[20] Alternatively, when mp is large, such as when p exceeds n, an initial screening procedure can be carried out to remove some irrelevant covariates, and the above procedures can be performed on the remaining covariates.

Given the group lasso estimator $\widehat{\boldsymbol{\beta}}$ (or the debiased estimator $\widetilde{\boldsymbol{\theta}}$), we define the values of $\widehat{\beta}_j(\cdot)$ outside the grid by linear interpolation. Then, the cumulative baseline hazard function $\Lambda(t) \equiv \int_0^t \lambda(s)\,\mathrm{d}s$ can be

estimated as follows:

$$\widehat{\Lambda}(t) = \int_0^t \frac{\sum_{i=1}^n \left\{ \mathrm{d}N_i(s) - Y_i(s)\widehat{\boldsymbol{\beta}}(W_i)^{\mathrm{T}} \boldsymbol{X}_i \, \mathrm{d}s \right\}}{\sum_{i=1}^n Y_i(s)},$$

which is essentially the Breslow estimator within the framework of the additive hazards model.[3]

3 Simulation studies

We assessed the prediction and variable selection performance of the proposed methods through simulation studies. We generated the components of W and $\boldsymbol{X}$ independently from $\mathrm{Uniform}(0, 1)$ and $\mathrm{Uniform}(-\sqrt{3}, \sqrt{3})$, respectively. The outcome variable was generated from the following hazard function:

$$\begin{aligned}\lambda(t \mid W = w, \boldsymbol{X}) &= \lambda(t) + \beta_1(w)X_1 + \beta_2(w)X_2 + \beta_3(w)X_3 + \beta_4(w)X_4 + \beta_5(w)X_5 \\ &= t + \{1 - \tanh(\pi w)\}X_1 + \{1 - \tanh(\pi(1-w))\}X_2 + 0.2X_3 + 0.2X_4 + 0.2X_5.\end{aligned}$$

The functions $\beta_1(w)$ and $\beta_2(w)$ are plotted in Figure 1. We set the censoring time to follow an exponential distribution with the mean chosen to yield a censoring rate of about 30%. We considered a sample size of $n = 500$ and $p = 5$, 20, 50, 100, and 200.

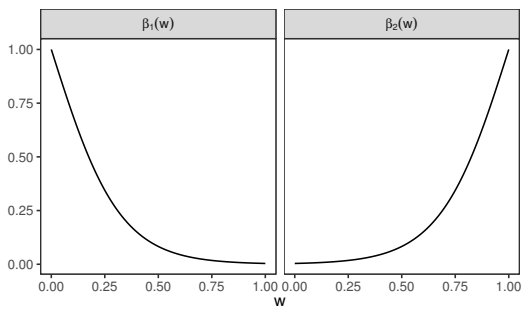

Figure 1. True varying coefficient functions in the simulation studies.

We compared three estimation approaches: the additive hazards model on the linear predictor $\boldsymbol{X}$, referred to as the "constant" method; the varying coefficient additive hazards model estimated using the local kernel method, referred to as the "local" method; and the varying coefficient additive hazards model estimated using the proposed method, referred to as the "global" method. For the local method, we impose a group lasso penalty[18] and estimate $\boldsymbol{\beta}$ over a grid $(w_1, \ldots, w_m)$ by minimizing

$$\sum_{k=1}^m \left\{ \frac{1}{2}\boldsymbol{\beta}(w_k)^{\mathrm{T}} \boldsymbol{V}(w_k, w_k)\boldsymbol{\beta}(w_k) - \boldsymbol{b}(w_k)^{\mathrm{T}}\boldsymbol{\beta}(w_k) \right\} + \lambda \sum_{j=1}^p \left\| \boldsymbol{\beta}_j \right\|_2.$$

Here $\boldsymbol{V}(w_k, w_k)$ and $\boldsymbol{b}(w_k)$ are equal to the corresponding quantities in the global method with $m = 1$. In the local kernel estimation method, each term in the first summation in the loss function does not depend

on other grid points. For the kernel methods, we considered a Gaussian kernel of $K(x) \propto \exp(-x^2/2)$. Following Silverman,[22] we set $h = (4/3n)^{1/5}\widehat{\sigma}$, where $\widehat{\sigma}$ is the empirical standard deviation of W. We obtained the estimates of $\boldsymbol{\beta}$ at an evenly spaced grid over $[0, 1]$ with size 6, and we used linear interpolation to estimate $\boldsymbol{\beta}$ outside the grid. For each approach, we considered conventional regression with $\lambda = 0$, referred to as "unpenalized," and penalized regression with λ selected by 5-fold cross-validation, referred to as "lasso." For the global method, we included the debiased estimator discussed in Section 2.3, using the model selected by lasso as basis, referred to as "debiased."

We evaluated the performance of each method in terms of variable selection and prediction. For variable selection, we reported the sensitivity and the false discovery rate. Sensitivity is the proportion of correctly identified signal variables among all true signal variables. False discovery rate is the proportion of noise variables that are incorrectly identified as signal variables among all selected variables. For prediction, we reported the mean-squared error, defined as $\mathrm{E}[\{\widehat{\boldsymbol{\beta}}(W) - \boldsymbol{\beta}(W)\}^{\mathrm{T}}\boldsymbol{X}]^2$. We computed the concordance index between $\widehat{\boldsymbol{\beta}}(W)^{\mathrm{T}}\boldsymbol{X}$ and the event time using an independent sample. The simulation results based on 500 replicates are summarized in Table 1. Also, we constructed the 95% confidence bands for $\beta_1(w), \ldots, \beta_5(w)$ over $w \in [0.2, 0.8]$ using the perturbation method discussed in Section 2. The estimated coefficients and the corresponding confidence bands are plotted in Figure 2. The estimated cumulative baseline hazard functions are plotted in Figure 3. The coverage rates of the 95% confidence bands are presented in Table 2.

Overall, the global estimator yields higher prediction accuracy than the alternatives. In all simulation settings, the proposed method yields superior prediction performance, with the lowest mean-squared error and highest concordance index values. The debiased estimators are generally more accurate. The local method also shows competitive performance but is generally inferior to the global method, probably because the proposed method accounts for the fact that all subjects share the same baseline hazard function. In terms of variable selection, the kernel methods have substantially lower FDR compared with the constant method, while maintaining similar SEN values. This indicates that model misspecification can lead to poor interpretation.

For $p = 50$, 100, and 200, the MSE values obtained from the unpenalized constant methods are lower than those obtained from the unpenalized kernel methods. It is important to note that incorporating irrelevant variables into a model introduces unnecessary complexity, which can result in overfitting. In such cases, kernel estimators tend to perform worse than constant estimators derived from the misspecified model. This emphasizes the importance of imposing penalization on varying coefficient models to mitigate the effects of overfitting and enhance their overall performance.

In the additional simulation studies presented in the supplemental material, we assessed the performance of the proposed method under non-monotone coefficient functions, misspecified outcome model, and different shapes of the baseline hazard function. The results are consistent with those from the previous simulations, suggesting that the global approach outperforms the alternatives.

4 Real data analysis

We applied the proposed method to identify aging-related genes and examine their effects on cancer progression. We used a dataset of patients with kidney renal clear cell carcinoma (KIRC) obtained from The Cancer Genome Atlas (TCGA),[23] available at `https://gdac.broadinstitute.org`. We investigated the effects of gene expressions on time to death since initial diagnosis, allowing the effects of

Table 1. Prediction and variable selection performance of different estimation methods in the simulation studies.

		Constant		Local		Global		
	p	**Unpenalized**	**Lasso**	**Unpenalized**	**Lasso**	**Unpenalized**	**Lasso**	**Debiased**
MSE	5	0.162	0.164	0.038	0.046	0.032	0.039	0.036
	20	0.174	0.177	0.095	0.145	0.090	0.135	0.082
	50	0.217	0.186	0.401	0.159	0.386	0.151	0.091
	100	0.337	0.208	4.899	0.167	4.596	0.157	0.096
	200	0.797	0.221	*	0.182	*	0.173	0.112
C-index	5	0.686	0.685	0.735	0.734	0.736	0.735	0.735
	20	0.662	0.667	0.708	0.701	0.709	0.705	0.717
	50	0.637	0.655	0.651	0.685	0.652	0.688	0.702
	100	0.631	0.674	0.596	0.689	0.597	0.691	0.714
	200	0.594	0.668	0.530	0.704	0.537	0.705	0.721
SEN	5		1.000		1.000		1.000	
	20		1.000		0.986		0.990	
	50		1.000		0.964		0.962	
	100		0.999		0.939		0.938	
	200		0.996		0.899		0.888	
FDR	5		0.000		0.000		0.000	
	20		0.500		0.091		0.101	
	50		0.610		0.036		0.032	
	100		0.670		0.027		0.023	
	200		0.684		0.018		0.012	

Abbreviations: MSE, mean-squared error; C-index, concordance index; SEN, sensitivity; FDR, false discovery rate.
* Omitted due to excessively large values in the case of moderately large number of variables.

gene expression to vary with age. The dataset used in our study consisted of 531 samples. We considered the overall survival as the outcome of interest. The median time to censoring or death is 3.29 years, and the censoring rate is 67.04%. After discarding genes with zero expressions for 50% or more subjects, the data set consists of 17,714 gene expressions. We set W as age at initial pathologic diagnosis, and we set $\boldsymbol{X}$ to consist of 300 gene expressions that exhibited the most significant marginal association with overall survival. In particular, we fitted an additive hazards model for each of the 17,714 gene expressions against overall survival and computed the p-values for the effects of individual gene expressions. We then selected the 300 gene expressions with the smallest p-values, indicating the most significant marginal associations with overall survival. We set an evenly-spaced grid points over [52,70] with size 5. We standardized the components of $\boldsymbol{X}$ to have zero mean and unit variance. We used 5-fold cross-validation to select λ. The selected gene expressions and their estimated coefficients and 95% confidence bands are presented in Figure 4. For comparison, we also performed lasso on the model with linear predictor $\boldsymbol{X}$, and the estimated coefficients are presented in Table 3.

The proposed method identified 6 gene expressions that are associated with overall survival. Among these, 4 gene expressions (CSTF3, FASN, ONECUT2, and ZNF117) were selected by the lasso on the model with linear predictor, as shown in Table 3. It is worth noting that most of the selected gene

Table 2. Coverage rates of 95% confidence bands for varying coefficient functions in the simulation studies.

p	w	β_1	β_2	β_3	β_4	β_5
5	0.2	0.982	0.994	0.992	0.984	0.996
	0.4	0.992	0.992	0.984	0.998	0.988
	0.6	0.990	0.990	0.982	0.990	0.986
	0.8	1.000	0.962	0.988	0.990	0.988
	Overall	0.964	0.938	0.952	0.962	0.964
20	0.2	0.978	0.988	0.994	0.982	0.982
	0.4	0.990	0.998	0.986	0.988	0.982
	0.6	0.992	0.996	0.992	0.988	0.990
	0.8	0.990	0.994	0.986	0.988	0.982
	Overall	0.952	0.978	0.960	0.948	0.936
50	0.2	0.976	0.994	0.994	0.982	0.978
	0.4	0.998	0.998	0.990	0.988	0.992
	0.6	0.982	0.992	0.998	0.998	0.984
	0.8	0.998	0.976	0.988	0.980	0.990
	Overall	0.954	0.960	0.970	0.950	0.952
100	0.2	0.970	0.998	0.982	0.990	0.984
	0.4	0.992	0.992	0.988	0.994	0.988
	0.6	0.992	0.984	0.984	0.992	0.996
	0.8	0.992	0.986	0.982	0.984	0.976
	Overall	0.946	0.960	0.936	0.960	0.948
200	0.2	0.970	0.998	0.980	0.992	0.986
	0.4	0.996	0.990	0.990	0.982	0.994
	0.6	0.990	0.992	0.992	0.986	0.998
	0.8	0.996	0.976	0.984	0.982	0.988
	Overall	0.952	0.956	0.946	0.948	0.970

expressions have been previously reported as relevant to KIRC.[24–27] Also, the expressions of CSTF3 and ZNF117 are known prognostic factors in other common cancer types.[28,29] Although the confidence bands for the estimated coefficients shown in Figure 4 do not clearly rule out the constant effect across the observed range, the effect of ONECUT2 tends to be larger in magnitude as age increases, suggesting that age appears to strengthen the effect of ONECUT2 on overall survival.

To assess the ability of the proposed method to distinguish between informative and redundant variables, and to evaluate its performance in a more challenging situation involving a higher dimension of $\boldsymbol{X}$, we introduced an additional 300 redundant covariates, which were generated from a standard normal distribution. We then applied the proposed method to select variables and repeated this procedure 100 times. Among the 100 replicates, two redundant variables were selected in 3 replicates, one redundant variable was selected in 7 replicates, and no redundant variables were selected in the remaining replicates. This indicates that the proposed method was able to effectively distinguish between informative and redundant variables in most cases, even in the presence of a substantial number of extraneous covariates. Taking the selected model based on the original data (with the 300 gene expressions selected by marginal screening) as the true model, the sensitivity of the proposed methods on the augmented data sets is 0.627 and the false discovery rate is 0.022. Table 4 shows the selection proportions of the selected gene expressions after adding the redundant variables in 100 replicates.

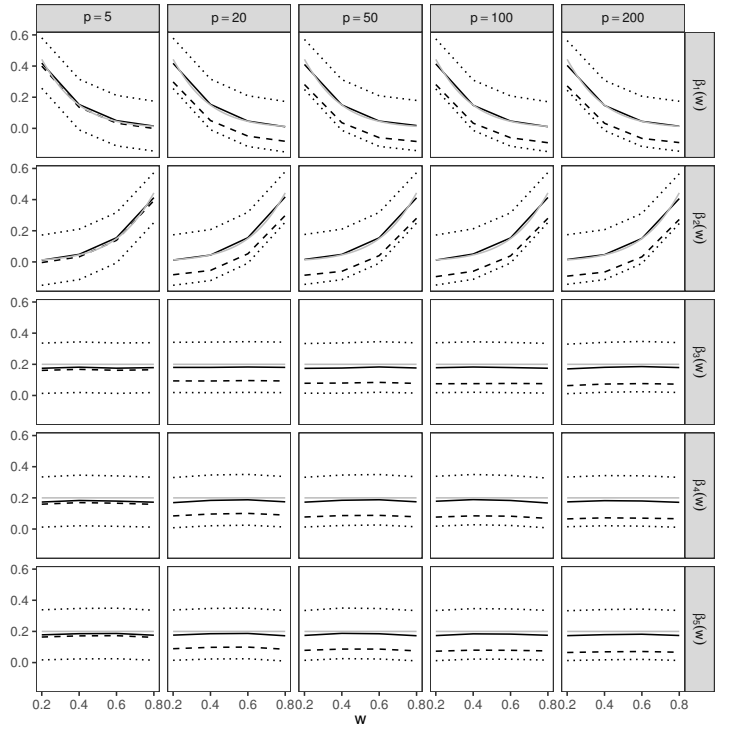

Figure 2. The estimated coefficient functions (dashed), the debiased estimates (solid in black), true coefficient functions (solid in gray), and 95% confidence bands (dotted) in the simulation studies.

5 Discussion

In this paper, we propose a penalized estimation method for the varying coefficient additive hazards model based on a global kernel estimator.[10] Our proposed method tackles the challenges arising from high-dimensional data and variable selection. Through extensive simulation studies, we have demonstrated that our method effectively identifies covariate effects and achieves accurate predictions. The proposed method can be widely applied to identify prognostic factors in many genomic studies of chronic diseases.

One advantage of the additive hazards model is its computational efficiency. In the state-of-the-art coordinate algorithm for the penalized Cox model,[30] there are two layers of iterations. The outer iteration repeatedly performs quadratic approximation on the log partial likelihood, and the inner iteration cycles through components of the regression parameters. By contrast, due to the simple form of the loss function under the additive hazards model, only the inner iteration is required, as the unpenalized loss function is already quadratic. This distinction underscores the computational benefits of our approach. In future research, we would extend the penalized estimation method to encompass more general models, including transformation models that incorporate the proportional hazards model as a special case.

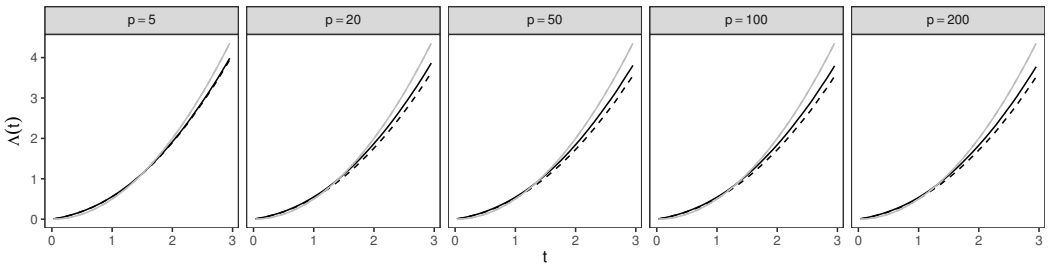

Figure 3. The estimated cumulative baseline hazard functions based on the lasso estimates (dashed), the estimated cumulative baseline hazard functions based on the debiased estimates (solid in black), and the true cumulative baseline hazard functions in the simulation studies.

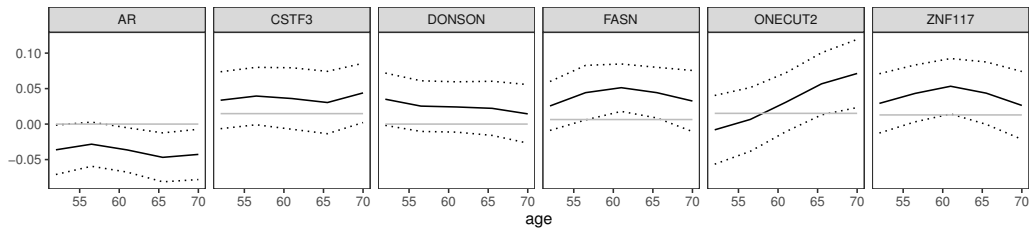

Figure 4. The debiased estimates (solid), the 95% confidence bands (dotted), and the lasso estimates from the constant method (gray) in the KIRC analysis. The constant method identified 28 additional gene expressions compared to the proposed method.

While the additive hazards model offers flexibility in modeling survival data and serves as an alternative to the commonly used proportional hazards model, there is a concern about potentially negative estimates for the hazard function. One approach to address this issue is to impose constraints on the model parameters during the estimation process, ensuring that the resulting hazard function remains positive.[31] Nevertheless, even without the constraint, negative estimated values of the hazard do not affect the estimation of the regression parameters. In all simulation studies and real data analyses conducted in

Table 3. Estimated coefficients of the additive hazards model with linear predictors in the KIRC analysis.

Gene expression	Coefficient	Gene expression	Coefficient
ISPD	−0.016	COL7A1	0.004
MSH3	−0.005	SOCS3	0.004
CYP3A7	−0.003	MBOAT7	0.005
PTPRG	−0.001	HES7	0.005
AUP1	0.000	RGS17	0.006
BREA2	0.000	PRH1	0.006
TMEM81	0.001	FASN	0.006
TRIM27	0.001	HOXA2	0.006
CCDC19	0.001	RGS20	0.007
CRABP2	0.001	RNF34	0.007
TNNT1	0.002	LOC100134259	0.010
SMARCD1	0.002	ZNF117	0.013
SEC61A2	0.002	CSTF3	0.015
LOC286467	0.002	ONECUT2	0.015
PYCR1	0.002	CILP	0.017
FOXA1	0.003	OTX1	0.019

Table 4. Frequency table of the selected variables over 100 replicates in the KIRC analysis with artificially augmented data.

Gene expression	AR	CSTF3	DONSON	FASN	ONECUT2	ZNF117
Frequency	94	7	100	35	100	40

this paper, the baseline hazard function estimates are positive, except occasionally at small time values where they may be negative.

There are several possible directions for future research. First, the proposed method can be extended to identify covariates with constant or varying effects on the survival outcome. In practical applications, certain covariates may have a constant effect on the hazard rate, but prior knowledge about which covariate effect on the hazard rate can be reasonably assumed to be constant is not always available. To identify the true model structure, we can impose separate penalization on the constant and varying covariate effects. Specifically, for $j = 1, \ldots, p$, we can decompose $\beta_j(W)$ into varying and constant components, such that $\beta_j(W) = \alpha_j^* + \beta_j^*(W)$, and the intercept of $\beta_j^*(\cdot)$ is fixed for model identifiability. Instead of (3), we minimize

$$\frac{1}{2}\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{V}\boldsymbol{\beta} - \boldsymbol{b}^{\mathrm{T}}\boldsymbol{\beta} + \lambda_1 \sum_{j=1}^{p} |\alpha_j^*| + \lambda_2 \sum_{j=1}^{p} \|\boldsymbol{\beta}_j^*\|_2,$$

where $\boldsymbol{\beta}_j^* = (\beta_j^*(w_1), \ldots, \beta_j^*(w_m))^{\mathrm{T}}$, and λ_1 and λ_2 are tuning parameters for the constant and varying parts, respectively. Under this loss function, $\widehat{\beta}_j^*$ could be shrunk to zero while $\widehat{\alpha}_j^*$ is nonzero, which represents a nonzero but constant effect of X_j. Although this approach yields a more interpretable model, it introduces computational challenges due to the additional penalty term and the need to search for tuning parameters in two dimensions.

The proposed method can also be extended to accommodate a multivariate $\boldsymbol{W}$. This extension would allow the kernel function to take in a multivariate vector and return a corresponding weight. However, it is important to note that a larger sample size is typically required to ensure reliable estimation when dealing with a higher dimensional $\boldsymbol{W}$. Additionally, we can allow the varying effect of each covariate in $\boldsymbol{X}$ to depend on different components of $\boldsymbol{W}$ by employing distinct kernel matrices for each component of $\boldsymbol{X}$. Furthermore, we can consider a more flexible interaction structure by replacing $X_j\beta_j(\boldsymbol{W})$ with $\beta_j(X_j, \boldsymbol{W})$. Nonetheless, this flexibility can complicate interpretation and may increase the risk of overfitting, particularly in studies with limited sample sizes.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Supplemental material

Supplemental materials for this article, which contain additional simulation results, are available online.

References

1. Cox DR. Regression models and life-tables. *J R Stat Soc Ser B Methodol* 1972; 34: 187–202.
2. Aalen OO. A linear regression model for the analysis of life times. *Stat Med* 1989; 8: 907–925.
3. Lin DY and Ying Z. Semiparametric analysis of the additive risk model. *Biometrika* 1994; 81: 61–71.
4. Thoroddsen A, Einarsson G, Hardarson S et al. Renal cell carcinoma in young compared to older patients: comparison of clinicopathological risk factors and survival. *Scand J Urol* 2008; 42: 121–125.
5. Yuan J, Duan F, Zhai W et al. An aging-related gene signature-based model for risk stratification and prognosis prediction in breast cancer. *Int J Women's Health* 2021; 13: 1053–1064.
6. Xu Q and Chen Y. An aging-related gene signature-based model for risk stratification and prognosis prediction in lung adenocarcinoma. *Front Cell Dev Biol* 2021; 9: 685379.
7. Yue T, Chen S, Zhu J et al. The aging-related risk signature in colorectal cancer. *Aging (Albany NY)* 2021; 13: 7330–7349.
8. Li J, Gui C, Yao H et al. An aging and senescence-related gene signature for prognosis prediction in clear cell renal cell carcinoma. *Front Genet* 2022; 13: 871088.
9. Spainhour JCG and Qiu P. Identification of gene-drug interactions that impact patient survival in TCGA. *BMC Bioinform* 2016; 17: 409.
10. Ng HM and Wong KY. A global kernel estimator for partially linear varying coefficient additive hazards models. *Lifetime Data Anal* 2025; 31: 205–232.
11. Ma S and Huang J. Lasso method for additive risk models with high dimensional covariates. Technical Report No. 347, Department of Statistics and Actuarial Science, The University of Iowa, https://stat.uiowa.edu/sites/stat.uiowa.edu/files/2024-04/LASSO-Method-for-Additive-Risk-Models-with-High-Dimensional-Covariates.pdf, 2005.
12. Leng C and Ma S. Path consistent model selection in additive risk model via Lasso. *Stat Med* 2007; 26: 3753–3770.
13. Martinussen T and Scheike TH. Covariate selection for the semiparametric additive risk model. *Scand J Stat* 2009; 36: 602–619.

14. Lin W and Lv J. High-dimensional sparse additive hazards regression. *J Am Stat Assoc* 2013; 108: 247–264.
15. Liu L, Liu Y, Su F et al. Variable selection and structure estimation for ultrahigh-dimensional additive hazards models. *Can J Stat* 2021; 49: 826–852.
16. Qu L, Song X and Sun L. Identification of local sparsity and variable selection for varying coefficient additive hazards models. *Comput Stat Data Anal* 2018; 125: 119–135.
17. Tibshirani R. Regression shrinkage and selection via the lasso. *J R Stat Soc Ser B Methodol* 1996; 58: 267–288.
18. Yuan M and Lin Y. Model selection and estimation in regression with grouped variables. *J R Stat Soc Ser B Methodol* 2006; 68: 49–67.
19. Friedman J, Hastie T and Tibshirani R. A note on the group lasso and a sparse group lasso, 2010. Retrieved from: https://arxiv.org/pdf/1001.0736.pdf.
20. Ning Y and Liu H. A general theory of hypothesis tests and confidence regions for sparse high dimensional models. *Ann Stat* 2017; 45: 158–195.
21. Tian L, Zucker D and Wei L. On the Cox model with time-varying regression coefficients. *J Am Stat Assoc* 2005; 100: 172–183.
22. Silverman BW. *Density Estimation for Statistics and Data Analysis*. CRC press, 1986.
23. The Cancer Genome Atlas Network. Comprehensive molecular characterization of clear cell renal cell carcinoma. *Nature* 2013; 499: 43–49.
24. Zhao Z, Liu Y, Liu Q et al. The mRNA expression signature and prognostic analysis of multiple fatty acid metabolic enzymes in clear cell renal cell carcinoma. *J Cancer* 2019; 10: 6599–6607.
25. Hu C, Fang D, Xu H et al. The androgen receptor expression and association with patient's survival in different cancers. *Genomics* 2020; 112: 1926–1940.
26. Klümper N, Blajan I, Schmidt D et al. Downstream neighbor of SON (DONSON) is associated with unfavorable survival across diverse cancers with oncogenic properties in clear cell renal cell carcinoma. *Transl Oncol* 2020; 13: 100844.
27. Roldán FL, Izquierdo L, Ingelmo-Torres M et al. Prognostic gene expression-based signature in clear-cell renal cell carcinoma. *Cancers* 2022; 14: 3754.
28. Zhang MH and Liu J. Cleavage stimulation factor 2 promotes malignant progression of liver hepatocellular carcinoma by activating phosphatidylinositol 3'-kinase/protein kinase B/mammalian target of rapamycin pathway. *Bioengineered* 2022; 13: 10047–10060.
29. Topno R, Singh I, Kumar M et al. Integrated bioinformatic analysis identifies UBE2Q1 as a potential prognostic marker for high grade serous ovarian cancer. *BMC Cancer* 2021; 21: 220.
30. Simon N, Friedman JH, Hastie T et al. Regularization paths for Cox's proportional hazards model via coordinate descent. *J Stat Softw* 2011; 39: 1–13.
31. Lu C, Goeman J and Putter H. Maximum likelihood estimation in the additive hazards model. *Biometrics* 2023; 79: 1646–1656.