Published in final edited form as:

Ann Appl Stat. 2021 September ; 15(3): 1291–1307. doi:10.1214/21-aos1446.

ESTROGEN RECEPTOR EXPRESSION ON BREAST CANCER PATIENTS' SURVIVAL UNDER SHAPE RESTRICTED COX REGRESSION MODEL

Jing Qin*, Geng Deng†, Jing Ning‡, Ao Yuan§, Yu Shen‡

*National Institution of Allergy and Infectious Diseases

†Wells Fargo

‡Department of Biostatistics, University of Texas MD Anderson Cancer Center

§Department of Biostatistics, Bioinformatics and Biomathematics, Georgetown University

Abstract

For certain subtypes of breast cancer, study findings show that their level of estrogen receptor expression is associated with their risk of cancer death, and also suggests a non-linear effect on the hazard of death. A flexible form of the proportional hazards model, $\lambda(t|x, z) = \lambda(t) \exp(z^T \beta) q(x)$, is desirable to facilitate a rich class of covariate effect on a survival outcome to provide meaningful insight, where the functional form of $q(x)$ is not specified except for its shape. Prior biologic knowledge on the shape of the underlying distribution of the covariate effect in regression models can be used to enhance statistical inference. Despite recent progress, major challenges remain for the semiparametric shape-restricted inference due to lack of practical and efficient computational algorithms to accomplish non-convex optimization. We propose an alternative algorithm to maximize the full log-likelihood with two sets of parameters iteratively under monotone constraints. The first set consists of the non-parametric estimation of the monotone-restricted function $q(x)$, while the second set includes estimating the baseline hazard function and other covariate coefficients. The iterative algorithm in conjunction with the pool-adjacent-violators algorithm makes the computation efficient and practical. The Jackknife resampling effectively reduces the estimator bias, when sample size is small. Simulations show that the proposed method can accurately capture the underlying shape of $q(x)$, and outperforms the estimators when $q(x)$ in the Cox model is mis-specified. We apply the method to model the effect of estrogen receptor on breast cancer patients' survival.

Keywords

Concave or Convex function; Cox proportional hazards model; Jackknife bias correction; Pool adjacent violators algorithm; Shape restricted inference

National Institution of Allergy Bethesda, MD 20817, USA, jingqin@niaid.nih.gov.

SUPPLEMENTARY MATERIAL

Supplement to "Estrogen Receptor Expression on Breast Cancer Patients' Survival under Shape Restricted Cox Regression Model"
(). Some theoretical proofs and additional analyses are relegated in the supplementary materials.

1. Introduction.

The management of breast cancer has made substantial progress, due to improvements in targeted systemic therapies. However, the pace of advances has lagged for certain subtypes, including triple-negative breast cancer, which is defined as estrogen receptor (ER)-negative, progesterone receptor-negative, and human epidermal growth factor 2 (HER2)-negative. Breast cancers with $< 1\%$ of ER expression are considered ER-negative, according to the American Society of Clinical Oncology/College of American Pathologists (11). However, some recent studies have suggested that patients with HER2-negative breast cancer have equivocal performance with ER expression between those with $< 1\%$ and those with 1-10% (Fujii et al., 2017; Yi et al., 2014). Given these uncertainties, it is desirable to use continuous ER expression data to assess how survival outcomes vary with the ER expression. The proportional hazards model has been the most commonly used regression model in survival analyses (Cox, 1975). One key assumption is that given covariate x , the conditional hazard $\lambda(t|x)$ is associated with the hazard of failure event as follows:

$$\lambda(t | x) = \lambda(t) \exp(x\alpha), \quad (1.1)$$

where $\lambda(t)$ is an unspecified baseline hazard and α is an unknown parameter. Using a large cohort data of breast cancer patients with HER2-negative tumors who received neo-adjuvant chemotherapy followed by definitive surgery between 1984 and 2013, Fujii et al. (2017) found that higher ER expression was associated with a lower risk of death. However, the proportional hazard assumption of ER expression as a continuous variable with a linear effect, or using either the cutoff value of 1% or 10% on the hazard function of death is significantly violated. The extent to which this is the case suggests that the Cox model with ER as a linear predictor may not capture its dynamic effect on survival. While it is biologically agreeable for the effect of ER expression on the hazard function to be monotonically decreasing, a more flexible semiparametric hazard model is desirable to enhance the interpretation of the ER effect by enforcing a flexible shape-restricted function of ER expression. Such a functional framework may facilitate a rich class of covariate effect to provide more meaningful clinical insights to assess its effect on the long-term overall survival.

Similar to the relationship between ER expression and hazard of death in breast cancer patients, it is common that some covariate measurement often satisfies certain shape-constrained functions in applications. We denote the observed data as (T, C, X, Z) , where T is a failure time, C is a censoring time, X is a continuous covariate (e.g., ER expression), and Z is a vector of other baseline covariates of interest. We consider a more general form of the conditional hazard function as

$$\lambda(t | x, z) = \lambda(t) \exp(z^T \beta) q(x), \quad (1.2)$$

where $q(x)$ is an unspecified, but shape-restricted function, which includes monotonic non-increasing, non-decreasing, or unimodal functions.

In recent statistical and econometric literature, shape-restricted inference has attracted considerable attention. For example, Cule, Samworth and Stewart (2010) and Doss and

Wellner (2016) discussed the maximum likelihood estimation of a multi-dimensional density under the log-concave and s-concave restrictions, Chen and Samworth (2016) studied generalized additive and index models with shape constraints, and Miratrix, Wager and Zubizarreta (2017) focused how to obtain an enhanced inference for the underlying distribution, mean, or other quantities by using information on the shape of a density, such as symmetry or log concave. A comprehensive review on shape-restricted nonparametric inferences can be found in the monograph by (Groeneboom and Jongbloed, 2014) and references therein. Above methods are mainly focused on uncensored data; discussions with right-censored data are limited.

For right-censored survival data, Chen et al. (2010) proposed a global partial likelihood as an alternative to the local partial likelihood of Tibshirani and Hastie (1987) to estimate a covariate effect under the proportional hazard model by replacing $\exp(x\alpha)$ in (1.1) with an unspecified function $q(x)$ without any shape-restriction. Although the global version of partial likelihood yields more efficient nonparametric estimators, the choice of the bandwidth for the kernel smoothing method is not addressed. When the hazard function is assumed to be a constant or piecewise constants in the Cox regression model, Ancukiewicz, Finkelstein and Schoenfeld (2003) have proposed a nonparametric maximum likelihood estimation if the covariate is a monotonic non-decreasing function on the hazard. Unfortunately, their numerical algorithms fail to converge to a local maximum in some cases. Flexible, fully Bayesian survival models with a shape-restricted covariate effect have been proposed in the literature, often assuming that the hazard function is piecewise linear in the Cox regression model (e.g. Murray et al. (2016), among others). Recently, Chung et al. (2018) proposed a pseudo-iterative convex minorant algorithm, which exploits the pool-adjacent-violators algorithm (PAVA) methods under model (1.2). The proposed computation algorithm is insightful, but the non-convergence of the algorithm is around 10% in practical settings, and the local consistency for the isotonic Cox model holds only when an initial guess of the regression function is sufficiently close to the true one. Despite the important progress, considerable practical challenges remain.

Motivated by the need of a general shape-constrained proportional hazards model with practical and robust computational algorithms, we propose a new iterative computational algorithm to fit the regression model and illustrate the effect of ER expression on the overall survival for breast cancer patients. The shape restriction may include non-decreasing, non-increasing, or unimodal shape-restrictions. Instead of maximizing the Cox partial likelihood subject to a shape restriction, we maximize the full likelihood using an iterative algorithm with two sets of unknown parameters. One set includes estimating nonparametric baseline hazard function and covariate parameters β , and the other set consists of parameters of the shape-restricted regression function. This iterative algorithm, also called a block coordinate descent algorithm (Bertsekas, 1999), has been used in the penalized regression problems, e.g., (Friedman, Hastie and Tibshirani, 2010), but not for right-censored data under the conventional semiparametric models. The proposed algorithm has a major computational advantage, as it uses existing software in SAS, R or Splus for the Cox model, with proper weights from the linear predictor in the function and paired with the well-known PAVA to enforce the shape-constraints.

Even though the maximum likelihood method leads to the most efficient estimators, in statistical literature, e.g., (Lancaster, 2000), it is well recognized that these estimates could have bias when the sample size is small and the number of nuisance parameters gets large, as in this context when both the baseline hazard function and $q(\cdot)$ are non-parametrically estimated. Typical bias-reduction approaches include conditioning the likelihood with nuisance parameters being cancelled, as in the partial likelihood approach in the Cox model or resampling bias correction methods (Quenouille, 1956; Efron, 1982). For the shape-restricted Cox model, the conditional likelihood method does not work. Nevertheless, the Jackknife correction serves the purpose well. Due to fast convergence of the proposed estimation algorithm, it is feasible to apply the Jackknife resampling method, which leads to considerable bias reduction in the estimators, and yields valid variance estimators.

This paper is organized as follows. In Section 2 we consider the case with one monotonic covariate, in addition to other baseline covariates under the proportional hazards regression model for data with and without right censoring. We propose a new alternative maximization approach with the monotone constraints under this flexible semiparametric model. Simulations are presented in Section 3. We illustrate the methods in detailed analyses using the motivating example in Section 4. We conclude this paper with discussions in Section 5. Some theoretical proofs and additional analyses are relegated in the supplementary materials (Qin et al., 2021).

2. Monotonic Shape-Restricted Covariate in Proportional Hazards Model.

2.1. Data without censoring.

We first consider failure times without censoring. The hazard function, including a monotone shape-constrained covariate effect in addition to other linear covariates $\mathbf{z}$, follows as

$$\lambda(t \mid x, \mathbf{z}) = \lambda(t) \exp(\mathbf{z}^T \boldsymbol{\beta}) / r(x), \quad (2.1)$$

where $r(\cdot)$ is an unspecified monotonic non-increasing or non-decreasing function, and $\mathbf{z}$ is a $p \times 1$ vector of additional covariates. Note that $r(\cdot)$ takes a reciprocal of $q(\cdot)$ on the hazard function in (1.2) for a reason, which will be made clear in Section 2.2. The corresponding probability density function is

$$f(t \mid x, \mathbf{z}) = \lambda(t) \exp(\mathbf{z}^T \boldsymbol{\beta}) / r(x) \exp\{-\Lambda(t) \exp(\mathbf{z}^T \boldsymbol{\beta}) / r(x)\},$$

where $\Lambda(t) = \int_0^t \lambda(u) du$ is the cumulative baseline hazard function.

Given observed data $(t_i, x_i, \mathbf{z}_i)$, $i = 1, 2, \dots, n$, the log of full likelihood is

$$\ell(r, \lambda, \boldsymbol{\beta}) = \sum_{i=1}^n [\log \lambda(t_i) + \mathbf{z}_i^T \boldsymbol{\beta} - \log r(x_i) - \Lambda(t_i) \exp(\mathbf{z}_i^T \boldsymbol{\beta}) / r(x_i)].$$

For notational simplicity, we assume $x_1 \leq x_2 \leq \dots \leq x_n$. In order to maximize this log-likelihood, we only need to estimate $\lambda(t)$ with jumps at failure data points, and $r(x)$ at x_i as defined by

$$\lambda_i = \lambda(t_i), \quad \Lambda(t_i) = \sum_{j=1}^n \lambda_j I(t_j \leq t_i), \quad r_i = r(x_i), \quad i = 1, 2, \dots, n.$$

The log-likelihood can be re-expressed with the discrete version as

$$\ell(\mathbf{r}, \boldsymbol{\lambda}, \boldsymbol{\beta}) = \sum_{i=1}^n \left[\log \lambda_i + \mathbf{z}_i^T \boldsymbol{\beta} - \log r_i - \sum_{j=1}^n \lambda_j I(t_j \leq t_i) \exp(\mathbf{z}_i^T \boldsymbol{\beta}) / r_i \right], \quad (2.2)$$

where $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_n)$ and $\mathbf{r} = (r_1, \dots, r_n)$. Since the functional forms of both $\lambda(t)$ and $r(x)$ are not specified, we only can identify them up to a multiplicative constant. Although we may fix $\lambda(\cdot)$ at t_1 or at another time point to be known, an alternative approach is to impose the constraint on the cumulative hazard function evaluated at the maximum observed data point Zhou (2015). Specifically, let $\Lambda(\max_{1 \leq i \leq n} t_i) = 1$. Thus, we have the following constraints to normalize $\lambda(t)$:

$$\lambda_i \geq 0, \quad \sum_{i=1}^n \lambda_i = 1. \quad (2.3)$$

In order to maximize this likelihood subject to the inequality and equality constraints on $r(\cdot)$ and $\lambda(\cdot)$, we need to introduce the Lagrange multipliers ν and ν_i , $i = 1, 2, \dots, n$. Let

$$H = - \sum_{i=1}^n \left[\log \lambda_i + \mathbf{z}_i^T \boldsymbol{\beta} - \log r_i - \sum_{j=1}^n \lambda_j I(t_j \leq t_i) \exp(\mathbf{z}_i^T \boldsymbol{\beta}) / r_i \right] + \nu h + \sum_{i=1}^n \nu_i g_i, \quad (2.4)$$

where the constraint equations are: $r_0 = 0$, $g_i = r_{i-1} - r_i$, $i = 1, 2, \dots, n$ and

$$h = \sum_{i=1}^n \lambda_i - 1.$$

Differentiating H with respect to unknown parameters, we have

$$\frac{\partial H}{\partial \lambda_k} = -\frac{1}{\lambda_k} + \sum_{j=1}^n I(t_j \geq t_k) \exp(\mathbf{z}_j^T \boldsymbol{\beta}) / r_j + \nu = 0,$$

$$\frac{\partial H}{\partial r_k} = \frac{1}{r_k} - \frac{\exp(z_k^T \boldsymbol{\beta})}{r_k^2} \sum_{j=1}^n I(t_j \leq t_k) \lambda_j - \nu_k + \nu_{k+1} = 0,$$

where $k = 1, \dots, n$, $\nu_k \geq 0$, $\sum_{k=1}^n \lambda_k - 1 = 0$, and

$$\nu_k(r_{k-1} - r_k) = 0, \quad 0 \leq r_1 \leq r_2 \leq \dots \leq r_n. \quad (2.5)$$

We can easily find

$$\lambda_k = \frac{1}{\sum_{j=1}^n I(t_j \geq t_k) \exp(z_j^T \boldsymbol{\beta}) / r_j + \nu},$$

where ν is determined by the constraint equation of $\sum_{k=1}^n \lambda_k = 1$, i.e.,

$$Q(\nu | \boldsymbol{\beta}, r) = \sum_{k=1}^n \frac{1}{\sum_{j=1}^n I(t_j \geq t_k) \exp(z_j^T \boldsymbol{\beta}) / r_j + \nu} - 1 = 0. \quad (2.6)$$

In the constraint equation (2.5), if $\nu_k > 0$, then $r_{k-1} = r_k$; this is called the inequality constraint active. If $\nu_k = 0$, however, this is called the inequality constraint inactive. In general it would be difficult to check analytically whether the inequality constraints are active or not, since there are 2^n possibilities for the solution. However, this is possible numerically by using the alternative maximization algorithm discussed in the next section. Moreover, to ensure the optimal solution, we have to check Karush-Kuhn-Tucker (KKT) conditions, which are the first-order necessary conditions for a solution in nonlinear programming to be optimal, provided that some regularity conditions are satisfied.

2.2. A maximization computation algorithm.

In this section we describe the step-by-step new iterative maximization algorithm to estimate the two sets of unknown parameters. Note that the log likelihood (2.2) is invariant by replacing $(\lambda_1, \dots, \lambda_n)$ and $(r_1, \dots, r_n)$ by $c \times (\lambda_1, \dots, \lambda_n)$ and $(r_1, \dots, r_n)/c$, where c is any positive constant. In another word, $\ell(c^{-1}r, c\lambda, \boldsymbol{\beta}) = \ell(r, \lambda, \boldsymbol{\beta})$. Thus, an anchor constraint is required for identifiability by imposing it on either r or λ . For numerical convenience, we have imposed the constraint on the hazard rate function λ . Specifically, the proposed algorithm enjoys a computational advantage by using existing software in SAS, R, or Splus for the Cox model with proper weights from the linear predictor in the function $r(\cdot)$ and paired with the well-known PAVA to enforce the shape-constraints.

Step 1. Given initial values $\Lambda^0(t_i) = \sum_{j=1}^n I(t_j \leq t_i) \lambda_j^0$ and $\boldsymbol{\beta}^0$, which are obtained from the Cox regression model $\lambda(dx, z) = \lambda(t) \exp(z^T \boldsymbol{\beta} + x\alpha)$, the contribution of the i -th subject in the likelihood (2.2) is

$$\frac{1}{r_i} \exp\{-\Lambda^0(t_i) \exp(z_i^T \boldsymbol{\beta}^0) / r_i\}.$$

Note that it has the form of an exponential density with unknown parameter r_i . For given $\boldsymbol{\beta}^0$ and $\lambda_i^0 > 0, i = 1, 2, \dots, n, i = 1, 2, \dots, n$, we can easily find the unrestricted maximum likelihood estimator (MLE) of $r_i, i = 1, 2, \dots, n$ as

$$\hat{r}_i^1 = \exp(z_i^T \boldsymbol{\beta}^0) \Lambda^0(t_i).$$

We use the monotonic, non-decreasing constraint on $r(\cdot)$ for illustration. Denote the updated estimators as $\hat{r}_1^1 \leq \dots \leq \hat{r}_n^1$. The result of PAVA creates index blocks $B_k, k = 1, \dots, m$ covering the full index set:

$$\bigcup_{k=1}^m B_k = \{1, 2, \dots, n\}$$

and $\hat{r}_i^1$ are constant within the same index block $B_k; \forall i \in B_k, \hat{r}_i^1 = \hat{r}_{B_k}^1$. Figure 1 shows the monotone shape-constraint estimator for $r(x)$ as a step function using PAVA. The point of using PAVA is to identify the active sets, within each active set the function is constant, and between active sets the function jumps. We denote $\hat{r}_{B_1}^1 < \hat{r}_{B_2}^1 < \dots < \hat{r}_{B_m}^1$ as strictly increasing values. PAVA yields the block-wise estimator as

$$\hat{r}_{B_k}^1 = \frac{\sum_{i \in B_k} \Lambda^0(t_i) \exp(z_i^T \boldsymbol{\beta}^0)}{n_{B_k}}, \quad (2.7)$$

where n_{B_k} is size of block B_k . As described in Robertson, Wright and Dykstra (1988), we can directly estimate the mean of each block with the monotone constraint by PAVA under model (2.1) with function $r(x)$, which takes the computational advantage of PAVA compared to the model specification with $q(\cdot)$ in (1.2). For fixed λ and $\boldsymbol{\beta}$ values, it is straightforward to prove that PAVA yields the shape-constrained MLE, which satisfies the KKT conditions specified in section 2.1.

Step 2. To obtain the MLE of $\boldsymbol{\beta}$ given $\hat{r}_1^1, \dots, \hat{r}_n^1$, we solve the following score equation, and denoted the estimator as $\hat{\boldsymbol{\beta}}^1$.

$$\sum_{i=1}^n z_i - \sum_{i=1}^n \sum_{j=1}^n \lambda_j^0 I(t_j \leq t_i) z_i \exp(z_i^T \boldsymbol{\beta}) / \hat{r}_i^1 = 0.$$

Step 3. Given $\hat{\beta}^1$, λ can be estimated by the normalized Breslow's estimator (Breslow, 1972), for $k = 1, 2, \dots, n$

$$\hat{\lambda}_k^1 = \frac{\delta_k \exp(z_k^T \hat{\beta}^1) / \hat{r}_k^1}{\sum_{j=1}^n I(t_k \leq t_j) \exp(z_j^T \hat{\beta}^1) / \hat{r}_j^1} \left/ \frac{\sum_{k=1}^n \delta_k \exp(z_k^T \hat{\beta}^1) / \hat{r}_k^1}{\sum_{j=1}^n I(t_k \leq t_j) \exp(z_j^T \hat{\beta}^1) / \hat{r}_j^1} \right|^{-1}.$$

Note that the normalized estimators satisfy the anchor constraint specified in (2.6) using the Lagrange method.

Step 4. Let $l = 1$, replace λ_i^0 , and $\hat{\beta}^0$ by $\hat{\lambda}_i^l$, $i = 1, 2, \dots, n$ and $\hat{\beta}^l$, respectively, and then go back to Step 1.

Step 5. Repeat Steps 1 to 4 multiple times until at the $(l + 1)$ th iteration, if

$$n^{-1} \sum_{i=1}^n |\hat{\lambda}_i^l - \hat{\lambda}_i^{l+1}| + n^{-1} \sum_{i=1}^n |\hat{r}_i^l - \hat{r}_i^{l+1}| + p^{-1} \sum_{j=1}^p |\hat{\beta}_j^l - \hat{\beta}_j^{l+1}| < \eta,$$

where $\beta = (\beta_1, \dots, \beta_p)$, and η is a pre-specified number (e.g. $\eta = 0.0001$).

Note that the negative full log-likelihood (2.4) is a convex function with respect to parameters λ and r , separately, but it may not be convex with respect to both sets of parameters jointly. This is a well known bi-convex optimization problem in the literature, for example, Gorski, Pfeuffer and Klamroth (2007). The alternative maximization algorithm described above for a general biconvex problem does not guarantee global optimality of the solution. Fortunately, the following Lemma and Proposition indeed indicate this algorithm converges to the optimal solution for the shaped-restricted Cox model.

Define $\ell(\lambda, r)$ and $\ell_p(r)$ as the log full likelihood and log partial likelihood, respectively. Following a derivation similar to Breslow (1972), we noted that the log of full likelihood becomes the log of partial likelihood under the Cox model, after profiling out λ in the full likelihood (i.e., $\ell(\hat{\lambda}(r), r) = \ell_p(r)$, where $\hat{\lambda}(r)$ satisfies the score equation $\ell(\lambda, r) / \lambda = 0$).

Lemma Assume that monotone function $r(x)$ is bounded away from 0 and ∞ , i.e., for any x , $0 < \epsilon \leq r(x) \leq M < \infty$, and $\lambda \geq 0$. If (λ_0, r_0) maximizes $\ell(\lambda, r)$, then r_0 maximizes $\ell_p(r)$, and $\ell(\lambda_0, r_0) = \ell_p(r_0)$. On the other hand if r_0 maximizes $\ell_p(r)$, then $(\hat{\lambda}(r_0), r_0)$ maximizes $\ell(\lambda, r)$.

Moreover, ℓ has an unique maximizer.

Proof: Suppose (λ_0, r_0) maximizes $\ell(\lambda, r)$, then, for any (λ, r) we have

$$\ell(\lambda_0, r_0) \geq \ell(\lambda, r).$$

In particular, for any λ , $\ell(\lambda_0, r_0) \geq \ell(\lambda, r_0)$. This implies

$$\ell(\lambda_0, \mathbf{r}_0) \geq \max_{\lambda} \ell(\lambda, \mathbf{r}_0) = \ell_P(\mathbf{r}_0). \quad (2.8)$$

On the other hand, for any fixed value of $\mathbf{r}$ including $\mathbf{r}_0$,

$$\ell(\lambda_0, \mathbf{r}_0) \leq \max_{\lambda} \ell(\lambda, \mathbf{r}_0) = \ell(\hat{\lambda}(\mathbf{r}_0), \mathbf{r}_0) = \ell_P(\mathbf{r}_0). \quad (2.9)$$

Combining the two inequalities from (2.8) and (2.9), we have

$$\ell(\lambda_0, \mathbf{r}_0) = \ell_P(\mathbf{r}_0).$$

If there are two sets of solutions, $(\lambda_0, \mathbf{r}_0)$ and $(\lambda_0^*, \mathbf{r}_0^*)$ such that

$$\max_{\lambda, \mathbf{r}} \ell(\lambda, \mathbf{r}) = \ell(\lambda_0, \mathbf{r}_0) = \ell(\lambda_0^*, \mathbf{r}_0^*),$$

then we have

$$\ell_P(\mathbf{r}_0) = \ell_P(\mathbf{r}_0^*).$$

Since $\ell(\mathbf{r})$ is strictly concave, it has a unique maximum point. Hence, $\mathbf{r}_0 = \mathbf{r}_0^*$. As a consequence, $\lambda_0^* = \lambda_0$, since $\ell(\lambda, \mathbf{r}_0)$ is strictly concave with respect to λ .

Next, if $\mathbf{r}_0$ maximizes $\ell(\mathbf{r})$, then for any $\mathbf{r}$, we have

$$\ell_P(\mathbf{r}_0) \geq \ell_P(\mathbf{r})$$

or

$$\ell(\hat{\lambda}(\mathbf{r}_0), \mathbf{r}_0) \geq \max_{\lambda} \ell(\lambda, \mathbf{r}) \geq \ell(\lambda, \mathbf{r})$$

for any $(\lambda, \mathbf{r})$. Similar arguments may be found in Murphy and Van der Vaart (2000)(p.449).

Proposition. Under the same assumption as in Lemma, the proposed iterative maximum likelihood algorithm produces a convergence sequence $(\mathbf{r}^k, \lambda^k)$, $k = 1, 2, \dots$, where $0 < \epsilon \leq r_1^k, r_2^k, \dots, r_n^k \leq M < \infty$, and its limiting value $\mathbf{r}^*$ maximizes the log partial likelihood.

Proof. We first consider model (2.1) without the linear term $Z^T \beta$. Chung et al. (2018) have shown that the log partial likelihood, $\ell(\mathbf{r})$ is a strictly concave function of $\mathbf{r}$ with the “anchor constraint” on $\mathbf{r}$. Thereby, after profiling λ in the full likelihood, the estimator of $\mathbf{r}$

maximizes the log partial likelihood of Chung et al. (2018). Given $\mathbf{r}^k$, the iterative maximization algorithm leads to an estimator of $\boldsymbol{\lambda}$, denoted by $\boldsymbol{\lambda}^{k+1} = \hat{\boldsymbol{\lambda}}(\mathbf{r}^k)$. Then, given $\boldsymbol{\lambda}^{k+1}$, $\mathbf{r}^{k+1}$ maximizes $\ell(\boldsymbol{\lambda}^{k+1}, \mathbf{r})$ with respect to $\mathbf{r}$, subject to the monotonic restriction. We can obtain following inequalities since $\ell(\hat{\boldsymbol{\lambda}}(\mathbf{r}^k), \mathbf{r}^k) = \ell(\boldsymbol{\lambda}^{k+1}, \mathbf{r}^k)$:

$$\ell_P(\mathbf{r}^k) = \ell(\boldsymbol{\lambda}^{k+1}, \mathbf{r}^k) < \ell(\boldsymbol{\lambda}^{k+1}, \mathbf{r}^{k+1}) < \ell(\hat{\boldsymbol{\lambda}}(\mathbf{r}^{k+1}), \mathbf{r}^{k+1}) = \ell_P(\mathbf{r}^{k+1}).$$

Therefore, $\ell(\mathbf{r}^k)$ is an increasing sequence of k . The proposed algorithm indeed falls into the class of “block coordinate descent methods” in Bertsekas (1999). For fixed $\mathbf{r}$, $\ell(\boldsymbol{\lambda}, \mathbf{r})$ is a strict concave function for $\boldsymbol{\lambda}$, and vice versa. Under the domain constraints $0 < \epsilon - r_1, r_2, \dots, r_n < M < \infty$, we can conclude that every limiting point from the alternative maximization algorithm is a stationary point (Proposition 2.7.1 in Bertsekas (1999)). Given above Lemma, the limiting point obtained from the algorithm also maximize the full likelihood.

Under model (2.1), with additional covariates term of $\exp(\mathbf{z}^T \boldsymbol{\beta})$, we can first fix $\boldsymbol{\beta}$ and then go through the proof as before. Since the iterative algorithm produces a continuous differentiable function of $\boldsymbol{\beta}$, we can find a convergence subsequence by maximizing $\boldsymbol{\beta}$ in the second stage in a compact space of $\boldsymbol{\beta}$.

Remark: The alternative maximum algorithm produces a sequence $\mathbf{r}^k$, such that the log partial likelihood $\ell(\mathbf{r}^k)$ is monotonically increasing. Since the log partial likelihood is strictly concave, every convergence sequence or every stationary point is the global maximum (Theorem 2.5, Nocedal and Wright (2006)). The proposition can be extended for the modified algorithm with right-censored data without much complication. The additional large sample properties and their proofs for the estimators are included in A.2 of Supplementary Material (Qin et al., 2021).

2.3. Data with right censoring.

In this section we consider the general setting, where survival data have right-censored observations. Denote the observed survival data as $\{t_i = \min(t_i^0, c_i), \delta_i = I(t_i^0 \leq c_i), x_i, \mathbf{z}_i, i = 1, 2, \dots, n\}$, where T^0 may be right censored by C . With right censoring, the log of likelihood is

$$\ell(\boldsymbol{\lambda}, \mathbf{r}, \boldsymbol{\beta}) = \sum_{i=1}^n [\delta_i \{\log \lambda_i + \mathbf{z}_i^T \boldsymbol{\beta} - \log r_i\} - \Lambda(t_i) \exp(\mathbf{z}_i^T \boldsymbol{\beta}) / r_i].$$

Without loss of generality, we let $x_1 \leq x_2 \leq \dots \leq x_n$ and $r_i = r(x_i)$, with the corresponding constraints of $r_1 \leq r_2 \leq \dots \leq r_n$.

Similar to the estimator for the conventional Cox model, the estimator of $\Lambda(t)$ jumps only at unique failure times. Assume there are n_1 unique failure time points, denoted as $t_{(1)}, \dots, t_{(n_1)}$; define the corresponding hazard function as $\lambda_{(i)} = \lambda(t_{(i)})$, $i = 1, \dots, n_1$. Similar to (2.4) for

data without right censoring, we maximize the following log-likelihood with the order constraints on $r(x)$,

$$H = \sum_{i=1}^n \left\{ \delta_i \left[\log \left\{ \sum_{j=1}^{n_1} I(t_i = t_{(j)}) \lambda_{(j)} \right\} + \mathbf{z}_i^T \boldsymbol{\beta} - \log r_i \right] - \Lambda(t_i) \exp(\mathbf{z}_i^T \boldsymbol{\beta}) / r_i \right\}, \quad (2.10)$$

subject to the monotone constraint, where $\Lambda(t) = \sum_{j=1}^{n_1} \lambda_{(j)} I(t_{(j)} \leq t)$, and the constraint equations are the same as (2.5). We describe the maximization algorithm for this more general setting without repeating the same steps described in Section 2.2, mainly emphasizing any deviation.

For given $\boldsymbol{\beta}^0$ and $\lambda_{(j)}^0 > 0, j = 1, 2, \dots, n_1$, the unrestricted maximum likelihood estimator of r_i for uncensored failure time (i.e., $\delta_i = 1$) is the same as before, $r_i^1 = \exp(\mathbf{z}_i^T \boldsymbol{\beta}^0) \Lambda^0(t_i)$. When t_i is right censored (i.e., $\delta_i = 0$), only $\Lambda^0(t_i) \exp(\mathbf{z}_i^T \boldsymbol{\beta}^0) / r_i$ remains in the log-likelihood. To maximize the log-likelihood (2.10), clearly the estimator of r_i should be as large as possible under the monotonic non-decreasing constraint $\hat{r}_1^1 \leq \dots \leq \hat{r}_n^1$. Thus, the estimator $\hat{r}_i^1 = \min\{\hat{r}_j^1 : j > i, \delta_j = 1\}$ serves the purpose. If the observed time corresponding to the largest value of x (i.e., x_n) is right censored, we assign $\hat{r}_n^1 = \max\{\hat{r}_j^1 : \delta_j = 1, j = 1, \dots, n\}$, the largest estimator associated with all failure data points. Using similar arguments in Proposition 1 of Chung et al. (2018), $\hat{r}$ jumps when moving from one active set to the next one, corresponding to the largest estimator of $\hat{r}$ associated with all uncensored failure time points. Finally, we will use the same method of PAVA to ensure the constraints on $r(x)$ as in (2.7).

2.4. Jackknife bias correction.

We present a Jackknife bias correction method for the monotone restricted MLE (Quenouille, 1956; Efron, 1982; Shao and Tu, 1995). When sample size is less than 100, we have found that such the resampling correction is essential to reduce the estimator bias. Let $\hat{\boldsymbol{\beta}}$ define the shape-constrained MLE obtained from Section 2 under the Cox model (2.1). Denote $\hat{\boldsymbol{\beta}}_{(-i)}$ as the shape-constrained MLE after the i -th observation being deleted. Then the bias corrected Jackknife estimator is

$$\hat{\boldsymbol{\beta}}_{jack} = n\hat{\boldsymbol{\beta}} - (n-1)\hat{\boldsymbol{\beta}}_{(\cdot)},$$

where $\hat{\boldsymbol{\beta}}_{(\cdot)}$ is the empirical average of the Jackknife replicates and equals to $\sum_{i=1}^n \hat{\boldsymbol{\beta}}_{(-i)} / n$. The corresponding Jackknife standard error is defined as

$$SE(\hat{\boldsymbol{\beta}})_{jack} = \left\{ \frac{n-1}{n} \sum_{i=1}^n (\hat{\boldsymbol{\beta}}_{(-i)} - \hat{\boldsymbol{\beta}}_{(\cdot)})^2 \right\}^{1/2}.$$

3. Simulation studies.

We conducted simulation studies to evaluate the finite sample performance of the proposed MLE algorithm under the shape-restricted proportional hazards model. We also compared the estimators with those of the standard Cox regression model under various functional assumption for $r(\cdot)$. We generated $x, z = (z_1, z_2)$ independently from $N(0, 1)$, $N(0, 1)$, and $Bernoulli(0.5)$, respectively. The survival times were generated from a Weibull distribution, with a shape parameter of 2 and a scale function of $\exp(z_1 + z_2)/r(x)$. Here we chose three different functions of $r(\cdot)$, representing three monotone relationships between the hazard function and covariate x : $r(x) = \exp(x)$ (Scenario 1), $r(x) = \exp(x^{1/3})$ (Scenario 2), and $r(x) = \exp(x^3)$ (Scenario 3). We considered three sample sizes, 100, 500, and 1000, with 1000 replicates for each scenario. We presented estimators of the proposed method, including the shape-restricted function $r(\cdot)$ and β , the estimators using the Jackknife bias-correction method, and the variance estimators (Shao and Wu, 1989). For comparison, we also estimated the corresponding covariate coefficients of x and z under the standard Cox regression model.

Table 1 summarizes the simulation results of the estimated β for uncensored data. The summary statistics are the empirical mean, empirical standard deviation (SD), average of the estimated standard error (ESE), and mean squared error (MSE), with and without the Jackknife bias correction. For the regression coefficients related to z_1 and z_2 in Scenario 1, the proposed method after the bias correction and the standard Cox regression performed well. As expected, the standard Cox regression outperformed the proposed method with smaller biases and variances, since all three covariates satisfy the Cox regression model, and the Cox model utilized all information effectively. On the other hand, the proposed method posed an additional nonparametric component ($r(x)$) on the hazard function for covariate x . For scenarios 2 and 3, the proposed method performed well in terms of the estimated regression coefficients β , especially with the Jackknife bias correction method, while the Cox regression had severely biased covariate coefficient estimators, due to mis-specification of the functional form on x .

As expected, the proposed shape-constrained model estimators had relatively large biases in the scenarios with a small sample size ($n=100$), although these biases decreased quickly with increasing sample sizes. This can be explained by the fact that $r(\cdot)$ function is unspecified, besides a monotonicity assumption. Also note that the estimators with the Jackknife method had considerable advantage to reduce biases comparing to their counterpart, without using the Jackknife method particularly for small sample size. Nevertheless, when the true model did not follow the proportional hazards assumption for covariate x (Scenarios 2 and 3), the estimation biases by the Cox regression were persistent across different sample sizes. For variance estimation, the Jackknife standard errors tended to overestimate the empirical variation, given small sample sizes, but were close to the empirical standard deviations for the scenarios with large sample sizes. In summary, the Jackknife method can capture the variability of the proposed estimating procedure, and correct biases under the proposed model.

Table 2 summarizes the simulation results for survival data with right censoring, in which the censoring times were generated independently from a Uniform distribution. The censoring percentages are set to be similar for scenarios 1-3, 22%, 22%, and 25%, respectively. Similar to our previous findings for uncensored data, the proposed method performed well under all scenarios, while the Cox regression produced substantial biases when $r(x) = \exp(x)$. As expected, the presence of censoring did not impact the estimation bias, but increased variation of the estimators. In our case, the standard deviation was increased around 10%, due to right-censored data.

Moreover, we estimated the functional forms of $r(x)$ under three scenarios with different sample sizes. Figure 2 plots the true underlying curve, and the empirical means of estimated curves $\log(r(x))$ under the Cox model and the proposed monotone restricted model. For Scenario 1, where the $r(x)$ satisfied the Cox model, both the proposed method and Cox regression model captured the underlying function (a linear function), although it is not surprising that the proposed method had large biases at the tails when sample sizes were small ($n=100$). With a sample size of 500 or 1000, the difference between the estimated curve by the proposed method and true curve were not distinguishable. For Scenarios 2 and 3, the underlying function $r(\cdot)$ exhibited more curvature than that assumed by the conventional Cox model. The proposed method can accurately capture the underlying shape of the function, and the estimation biases at the tail also decreased with an increasing sample size. In contrast, the estimated curves by the Cox regression deviated from the underlying function substantially, and the differences between the estimated and true curves did not have a decreasing trend with increasing sample sizes. Figure 3 shows the standardized cumulative baseline hazard function at the upper support point of the censoring distribution (i.e., $\Lambda(t) = t^2/25$), and the empirical means of estimated cumulative baseline hazard functions under the two models. When the function $r(x) = \exp(x)$ (Scenario 1), the performance of the proposed method was similar to those under the Cox model. When the function $r(x)$ deviated from the proportional hazards assumption (Scenarios 2 and 3), the estimated cumulative hazard functions under the shape-restricted model were much closer to the true underlying function, with decreasing estimation biases with increasing sample sizes.

4. Application.

The ER expression is typically measured by immunohistochemistry (IHC) assay, and quantified as percentage of tumor cells with positive nuclear staining with an IHC assay for breast cancer patients. It has been an ongoing debate on how to define ER positivity (Iwamoto et al., 2012); historically, many investigators consider 10% or greater as the threshold for defining ER-positive status, whereas 1% was recommended by recent joint guidelines by ASCO for all subtypes of breast cancer and the American College of Pathologists. However, it is uncertain if the same cutoff is proper for patients with HER2 negative breast cancer (Fujii et al., 2017). ER expression is strongly associated with patients' clinical prognosis and overall survival (OS), and how ER positivity is defined directly impacts the decision for use of estrogen-targeted therapy. Due to the lack of the clinical consensus regarding a meaningful cutoff point for ER expression among patients with HER2-negative tumors (Yi et al., 2014), we aimed to comprehensively evaluate the effect of the ER expression on OS among the HER2 negative patients cohort. We considered such a

cohort of patients, who received neoadjuvant chemotherapy followed by definitive surgery between 1982 and 2013 at The University of Texas MD Anderson Cancer Center. After excluding patients who had missing information regarding ER expression and survival information, we included a total of 3033 patients in the analysis. In our preliminary analysis, we fitted the Cox model by including ER group indicators with cutoffs of 1% and 10% to the continuous ER expression value as covariates. These cut-points have been conventionally used in breast cancer literature to categorize patients by their ER expression levels. The corresponding p-value for testing the proportional hazards (PH) assumption on ER expression is less than 0.0001, whereas age at diagnosis and pCR status satisfy the proportional hazards assumption (see Table 3). As expected, the corresponding continuous ER expression variable did not satisfy the proportional hazards model either (Grambsch and Therneau, 1994). While the cutoff of ER expression has been controversial in literature, the oncology community has agreed that the effect of ER expression on the hazard function monotonically decreases. For patients with neoadjuvant chemotherapies, an important predictor to long-term patient performance is whether they have achieved pathological complete response (pCR), defined as no invasive carcinoma in the breast or no tumor in the axilla at the time of surgery. Among the patients cohort, 427 (14%) had achieved pCR. The age of diagnosis ranged from 18 to 83 years with a median of 49 years. In addition to age and pCR status in the proportional hazards model, we applied the monotone shape-restricted function form on ER expression $r(x)$, to analyze the effect of ER expression as a continuous variable, as follows $\lambda(t|x, z) = \lambda(t) \exp(\beta_1 \text{age} + \beta_2 \text{pCR})/r(x)$. The median follow-up was 3.9 years for the OS (up to 14 years), and 659 patients died during the follow-up.

We analyzed the data under both the standard Cox regression model and the proposed monotonically non-increasing constraint on ER expression. Both models confirmed that HER2-negative breast cancer patients achieving pCR had substantially improved OS compared to those with residual disease, while the age of diagnosis did not significantly impact OS after adjusting the status of pCR and ER expression. Figure 4 shows the estimated shape-constrained nonparametric function for the effect of ER expression on OS and cumulative baseline hazard functions by both methods. For a fair comparison, we applied the normalized constraint to the estimated baseline cumulative hazard function under the standard Cox model, as well. The fitted curve by the shape-restricted model was a convex function and considerably deviated from the straight line representing the linear effect of ER expression from the standard Cox model. On the scale of $-\log(r(x))$, the effect of ER expression on the hazard of death decreased sharply with the increasing ER expression from 0 to 20%, and then decreased stably when the ER expression increased from 20%. In contrast to the functional form of ER expression on the survival outcome, the baseline cumulated hazards functions estimated under the two Cox models showed similar patterns. While the convex function confirms the violation of proportionality assumption of ER expression on OS, the descriptive plot may help investigators to better identify a proper cutoff point for defining ER positivity, as well. The shape-constrained nonparametric estimator for $r(\cdot)$ implies a cutoff of 20% for ER expression may be proper for HER2-negative patients in defining the positivity of ER.

Per reviewers' suggestions, we used the bootstrap approach to test for the covariate effect of ER expression on the outcome. Specifically, based on the estimated curve for the ER expression in Figure 4 of the paper, we used the area between the curve from the proposed model and the x-axis, $\hat{D} = \int_0^{100} |\log\{\hat{r}(x)\}| dx$, where $\hat{r}(x)$ is the estimated ER effect from the monotone shape-restricted Cox model. We estimated the standard error of $\hat{D}$ by using 1000 bootstrap replicates. The estimated area under the curve was 32.56 with the bootstrap standard error of 5.24, suggesting that ER has a non-zero effect on the long-term survival. We also used a heuristic graphical way to show the monotonicity of the shape-restricted Cox regression model (in Figure S1 of Supplementary Material) (Qin et al., 2021).

5. Discussion.

As demonstrated in the motivating breast cancer research project, the conventional proportional hazards model may not explain certain continuous variables well; a sufficient flexible functional form on such a variable is needed. On the other hand, prior biologic knowledge may suggest a shape-constraint functional form on the hazard function in regression models, which can be used to enhance statistical inference. However, shape-restricted inference with practical and efficient computational algorithms has been limited in survival analyses to perform the non-convex (or non-concave) optimization. With similar motivations, Chung et al. (2018) proposed an elegant partial likelihood estimation method by considering a pseudo-iterative convex minorant algorithm together with PAVA. While the algorithm is theoretically justified, the finite sample performance of the parameter estimation of β in model (1.2), and potential loss of efficiency for the estimators, were not reported when compared with the estimators under the conventional Cox model. Moreover, the pseudo-iterative convex minorant algorithm used a quadratic approximation to the Cox's partial likelihood in the Taylor expansion, with respect to $q(x)$. Their method ensures to find a sequence q^k increasing with k for the quadratic term, but not necessarily the log of partial likelihood. Conceptually, for an initial value not in an n^{-1} neighborhood of true value for $q(x)$, its quadratic approximation may not be even close to the partial likelihood.

In this paper, we propose an alternative algorithm to maximize the full log-likelihood with two sets of parameters, iteratively; the first set consists of the estimation of the shape-restricted nonparametric function $r(x)$, and the second set includes estimation of the baseline hazard function and the other covariate coefficient in the regression model. The proposition ensures that we can find a sequence of r^k to maximize the partial likelihood, since the log partial likelihood is strictly concave with respect to r , the convergence is guaranteed and our algorithm achieves 100% convergence rate. Another advantage of the algorithm is that the underlying exponential family structure is retained at the first step, when estimating $r(x)$. As a result, the well known PAVA algorithm can be applied directly to accomplish the order-restricted estimation for the regression function $r(x)$, when the baseline hazard and β are fixed. For fixed values on $r(x)$, we can then consider the estimated $r(\cdot)$ in the linear predictor (as a weight) to estimate the baseline hazard function and covariate coefficients β . We can use the "offset" option in R/Spplus directly to estimate the baseline hazard function and β , with a small modification to enforce the identifiability constraint. The iterative algorithm,

together with the existing PAVA algorithm, makes the computation extremely practical and efficient.

As indicated in our simulations, the shape-restricted maximum likelihood estimators under the proportional hazards model could have a large bias when the sample size is small. This is a common problem in likelihood-based inference, when the number of nuisance parameters is large (Neyman and Scott, 1948; Lancaster, 2000). After applying the Jackknife method in the estimation algorithm, the bias of the estimator is considerably reduced, as demonstrated in the simulation. The numerical results also show that estimators from the proposed shape-constraint model outperform the estimators from the conventional Cox model, when $\pi(x)$ is mis-specified. With some modification, the proposed method also can be extended to handle time-dependent covariate problems, subject to shape restrictions.

In this paper, we focused on the computation algorithm for the monotonic-constrained function on the hazard function. When $\pi(x)$ is a unimodal function a priori, the algorithm will become more complex. First, we need to identify the mode (or change point) of $\pi(x)$ by trying all possible x_k , $k = 1, 2, \dots, n$. Given the estimated mode at x_k^* , $\pi(x)$ can be similarly estimated under a non-decreasing constraint and a non-increasing constraint, respectively. In nonlinear programming literature, it has been well known that the convexity of a targeted function is the key for convergence of any computation algorithms. As pointed out in Proposition, the negative log-likelihood function (2.4) may not be convex jointly with respect to both λ and r in general.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements.

The research is partially supported by grant R01 CA193878 (JN, YS). The authors wish to thank Dr. Naoto Ueno and his group to share breast cancer patient cohort data, which motivated this research. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

References.

- Ancukiewicz M, Finkelstein DM and Schoenfeld DA (2003). Modelling the Relationship Between Continuous Covariates and Clinical Events Using Isotonic Regression. *Statistics in Medicine* 22 3151–3159. [PubMed: 14518020]
- Bertsekas DP (1999). *Nonlinear Programming*. Athena scientific Belmont.
- Breslow N (1972). Discussion of the paper by D. R. Cox. *Journal of the Royal Statistical Society. Ser. B* 34 216–217.
- Chen Y and Samworth RJ (2016). Generalized additive and index models with shape constraints. *Journal of the Royal Statistical Society: Ser. B* 78 729–754.
- Chen K, Guo S, Sun L and Wang J-L (2010). Global Partial Likelihood for Nonparametric Proportional Hazards Models. *Journal of the American Statistical Association* 105 750–760. [PubMed: 22844168]
- Chung Y, Ivanova A, Hudgens MG and Fine JP (2018). Partial Likelihood Estimation of Isotonic Proportional Hazards Models. *Biometrika* 105 133–148. [PubMed: 29808076]
- Cox DR (1975). Partial likelihood. *Biometrika* 62 269–276.

- Cule M, Samworth R and Stewart M (2010). Maximum Likelihood Estimation of a Multi-dimensional Log-concave Density. *Journal of the Royal Statistical Society: Ser. B* 72 545–607.
- Doss CR and Wellner JA (2016). Global rates of convergence of the MLEs of log-concave and s-concave densities. *Annals of Statistics* 44 954. [PubMed: 28966409]
- Efron B (1982). The jackknife, the bootstrap, and other resampling plans 38. *Siam*.
- Friedman J, Hastie T and Tibshirani R (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software* 33 1–22. [PubMed: 20808728]
- Fujii T, Kogawa T, Dong W, Sahin A, Moulder S, Litton J, Tripathy D, Iwamoto T, Hunt K, Pusztai L et al. (2017). Revisiting the Definition of Estrogen Receptor Positivity in HER2-negative Primary Breast Cancer. *Annals of Oncology* 28 2420–2428. [PubMed: 28961844]
- Gorski J, Pfeuffer F and Klamroth K (2007). Biconvex sets and optimization with biconvex functions: a survey and extensions. *Mathematical Methods of Operations Research* 66 373–407.
- Grambsch PM and Therneau TM (1994). Proportional Hazards Tests and Diagnostics Based on Weighted Residuals. *Biometrika* 81 515–526.
- Groeneboom P and Jongbloed G (2014). *Nonparametric Estimation Under Shape Constraints*. Cambridge University Press.
- Iwamoto T, Booser D, Valero V, Murray JL, Koenig K, Esteva FJ, Ueno NT, Zhang J, Shi W, Qi Y et al. (2012). Estrogen receptor (ER) mRNA and ER-related gene expression in breast cancers that are 1% to 10% ER-positive by immunohistochemistry. *J Clin Oncol* 30 729–734. [PubMed: 22291085]
- Lancaster T (2000). The incidental parameter problem since 1948. *Journal of Econometrics* 95 391–413.
- Miratrix LW, Wager S and Zubizarreta JR (2017). Shape-constrained Partial Identification of a Population Mean Under Unknown Probabilities of Sample Selection. *Biometrika* 105 103–114.
- Murphy SA and Van der Vaart AW (2000). On profile likelihood. *Journal of the American Statistical Association* 95 449–465.
- Murray TA, Hobbs BP, Sargent DJ and Carlin BP (2016). Flexible Bayesian Survival Modeling with Semiparametric Time-dependent and Shape-restricted Covariate Effects. *Bayesian Analysis* 11 381. [PubMed: 27042243]
- Neyman J and Scott EL (1948). Consistent Estimates Based on Partially Consistent Observations. *Econometrica* 16 1–32.
- Nocedal J and Wright SJ (2006). *Nonlinear Equations*. Springer.
- Qin J, Deng G, Ning J and Shen Y (2021). Supplement to “Estrogen Receptor Expression on Breast Cancer Patients’ Survival under Shape Restricted Cox Regression Model”. *AOAS2021SUPP*.
- Quenouille MH (1956). Notes on Bias in Estimation. *Biometrika* 43 353–360.
- Robertson T, Wright F and Dykstra R (1988). *Order Restricted Statistical Inference*. New York: Wiley.
- Shao J and Tu D (1995). *The Jackknife and Bootstrap*.
- Shao J and Wu CJ (1989). A General Theory for Jackknife Variance Estimation. *The Annals of Statistics* 17 1176–1197.
- Tibshirani R and Hastie T (1987). Local Likelihood Estimation. *Journal of the American Statistical Association* 82 559–567.
- Yi M, Huo L, Koenig K, Mittendorf E, Meric-Bernstam F, Kuerer H, Bedrosian I, Buzdar A, Symmans W, Crow J et al. (2014). Which Threshold for ER Positivity? A Retrospective Study Based on 9639 Patients. *Annals of Oncology* 25 1004–1011. [PubMed: 24562447]
- Zhou M (2015). *Empirical Likelihood Method in Survival Analysis*. Chapman and Hall/CRC.

Fig 1.
Illustration of PAVA solutions with block-wise jumps

Fig 2.

The empirical means of estimated curves $-\log\{r(x)\}$ under three scenarios (Scenario 1: $r(x) = \exp(x)$; Scenario 2: $r(x) = \exp(x^{1/3})$; Scenario 3: $r(x) = \exp(x^3)$).

Fig 3.

The empirical means of estimated cumulative baseline hazard functions under three scenarios (Scenario 1: $r(x) = \exp(x)$; Scenario 2: $r(x) = \exp(x^{1/3})$; Scenario 3: $r(x) = \exp(x^3)$).

Fig 4.
Estimated, Effects of ER on Overall Survival and Cumulative Hazard Functions.

Table 1

Summary of simulation studies for uncensored data

n	r(x)		Proposed shape restricted model			Cox Model		
			Jackknife Method					
			Bias (SD)	MSE	Bias (SD)	ESE	MSE	Bias (SD)
100	exp(x)	$\beta_1 = 2$	.451 (.291)	.288	.100 (.274)	.339	.085	.037 (.208)
		$\beta_2 = 2$	.463 (.385)	.363	.103 (.390)	.446	.163	.043 (.283)
	exp($x^{1/3}$)	$\beta_1 = 2$	.327 (.258)	.173	.063 (.245)	.302	.064	-.326 (.169)
		$\beta_2 = 2$	.341 (.364)	.249	.072 (.357)	.400	.133	-.321 (.276)
	exp(x^3)	$\beta_1 = 2$	.471 (.312)	.319	.169 (.308)	.369	.123	-.791 (.182)
		$\beta_2 = 2$	.490 (.429)	.424	.200 (.437)	.483	.231	-.777 (.280)
500	exp(x)	$\beta_1 = 2$	.153 (.099)	.033	.051 (.096)	.103	.012	.003 (.086)
		$\beta_2 = 2$	.157 (.125)	.040	.054 (.125)	.139	.018	.007 (.112)
	exp($x^{1/3}$)	$\beta_1 = 2$	.104 (.094)	.020	.030 (.091)	.098	.009	-.360 (.077)
		$\beta_2 = 2$	.109 (.122)	.027	.038 (.122)	.132	.016	-.363 (.115)
	exp(x^3)	$\beta_1 = 2$	.209 (.109)	.056	.107 (.120)	.113	.026	-.878 (.084)
		$\beta_2 = 2$	.212 (.134)	.063	.109 (.142)	.151	.032	-.858 (.114)
1000	exp(x)	$\beta_1 = 2$	.099 (.063)	.014	.034 (.063)	.067	.005	.001 (.057)
		$\beta_2 = 2$	.097 (.087)	.017	.031 (.087)	.092	.009	.000 (.079)
	exp($x^{1/3}$)	$\beta_1 = 2$	.067 (.062)	.008	.020 (.062)	.065	.004	-.370 (.053)
		$\beta_2 = 2$	.065 (.085)	.011	.019 (.086)	.089	.008	-.373 (.080)
	exp(x^3)	$\beta_1 = 2$	.150 (.069)	.027	.069 (.068)	.072	.009	-.889 (.060)
		$\beta_2 = 2$	.148 (.088)	.030	.069 (.089)	.098	.013	-.872 (.079)

ESE: estimated standard error; MSE: mean squared error; SD: standard deviation

Table 2

Summary of simulation studies for right-censored data.

n	$r(x)$		Proposed shape restricted model					Cox Model	
						Jackknife Method			
			Bias (SD)	MSE	Bias (SD)	ESE	MSE	Bias (SD)	MSE
100	exp(x)	$\beta_1 = 2$	.500 (.325)	.356	.123 (.304)	.387	.108	.064 (.228)	.056
		$\beta_2 = 2$	.481 (.445)	.429	.107 (.436)	.505	.201	.053 (.327)	.110
	exp($x^{1/3}$)	$\beta_1 = 2$	.380 (.300)	.235	.086 (.290)	.346	.091	-.287 (.205)	.125
		$\beta_2 = 2$	.375 (.404)	.304	.094 (.392)	.454	.162	-.284 (.307)	.175
	exp(x^3)	$\beta_1 = 2$	.435 (.361)	.319	.194 (.352)	.407	.161	-.575 (.207)	.374
		$\beta_2 = 2$	.421 (.480)	.408	.178 (.494)	.529	.276	-.570 (.326)	.431
500	exp(x)	$\beta_1 = 2$	.158 (.109)	.037	.055 (.107)	.113	.014	.008 (.096)	.009
		$\beta_2 = 2$	.159 (.145)	.046	.056 (.147)	.154	.025	.006 (.128)	.016
	exp($x^{1/3}$)	$\beta_1 = 2$	.114 (.101)	.023	.036 (.098)	.107	.011	-.351 (.083)	.130
		$\beta_2 = 2$	.117 (.141)	.034	.043 (.140)	.147	.021	-.355 (.132)	.144
	exp(x^3)	$\beta_1 = 2$	.183 (.120)	.048	.089 (.120)	.122	.022	-.650 (.086)	.429
		$\beta_2 = 2$	.187 (.154)	.059	.095 (.156)	.165	.033	-.632 (.126)	.415
1000	exp(x)	$\beta_1 = 2$	.097 (.071)	.014	.031 (.069)	.074	.006	.000 (.065)	.004
		$\beta_2 = 2$	.098 (.099)	.019	.032 (.098)	.102	.011	.000 (.091)	.008
	exp($x^{1/3}$)	$\beta_1 = 2$	.068 (.068)	.009	.020 (.067)	.071	.005	-.359 (.062)	.133
		$\beta_2 = 2$	.068 (.096)	.014	.019 (.096)	.098	.010	-.358 (.092)	.137
	exp(x^3)	$\beta_1 = 2$	.124 (.076)	.021	.053 (.074)	.079	.008	-.669 (.063)	.452
		$\beta_2 = 2$	.125 (.098)	.025	.056 (.097)	.108	.013	-.647 (.088)	.427

ESE: estimated standard error; MSE: mean squared error; SD: standard deviation

Table 3

Estimated regression coefficients with standard errors (SEs) from proposed method and the standard Cox regression.

	Proposed Method		Cox Model		P-value for PH
	Coeff.	SE	Coeff.	SE	
Age at diagnosis	−0.0037	0.0040	−0.0032	0.0037	0.929
pCR (Yes vs No)	−1.9971	0.2220	−1.9532	0.2140	0.101
ER precentage	NA	NA	−0.0140	0.0009	<0.000

The PH assumption is tested based on weighted residuals (Grambsch and Therneau, 1994).