Least squares type estimation for Cox regression model and specification error

P.L.Gradowska[a,*], R.M.Cooke[b]

[a]*Department of Applied Mathematics, Delft University of Technology, The Netherlands*
[b]*Department of Applied Mathematics, Delft University of Technology, The Netherlands, and Resources for the Future, Washington, DC*

Abstract

The manuscript introduces a new estimation procedure for the Cox proportional hazards model. The method proposed employs the sample covariance matrix of model covariates and alternates between estimating the baseline cumulative hazard function and estimating model coefficients. It is shown that the estimating equation for model parameters resembles the least squares estimate in a linear regression model, where the outcome variable is the transformed event time. As a result an explicit expression for the difference in the parameter estimates between nested models can be derived. Nesting occurs when the covariates of one model are a subset of the covariates of the other. The new method applies mainly to the uncensored data, but its extension to the right censored observations is also proposed.

Keywords: Cox regression, estimation, model specification, simulation, specification error

1. Introduction

The Cox (1972) proportional hazards model is the most commonly applied semiparametric regression model for assessing the effect of covariates

*Corresponding address: Department of Applied Mathematics, Delft University of Technology, Mekelweg 4, 2628 CD Delft, The Netherlands, Tel: +31 15 2783997, Fax: +31 15 2787255

Email addresses: p.l.gradowska@tudelft.nl (P.L.Gradowska), r.m.cooke@ewi.tudelft.nl (R.M.Cooke)

on the hazard ratio using censored survival data. It postulates that the conditional hazard function of an event time given covariates is a product of an unspecified baseline hazard function and an exponential function of a linear combination of covariates with unknown coefficients. The standard technique for estimating the regression parameters of this model is the partial likelihood method (Cox, 1972, 1975). The method has become popular mainly because it allows making inferences about the regression coefficients without involving the baseline hazard, yields an estimator with good theoretical properties and is programmed in the most common statistical software packages. The inference on the baseline hazard and the baseline cumulative hazard function is typically based on either the Breslow (1972, 1974) estimator or the Kalbfleisch and Prentice (1973) estimator. Both these estimators are in reasonable agreement (Lawless, 1982). Nevertheless, since the Breslow estimator is simpler to evaluate it is being used more frequently in practice.

Breslow (1972) demonstrated that the Cox's maximum partial likelihood estimator and the corresponding Breslow estimator for the baseline cumulative hazard function are identical to the estimators obtained by maximizing the full likelihood function for the Cox model with respect to the regression coefficients and the baseline hazard simultaneously assuming that the baseline hazard is piecewise constant between the uncensored observations and that all censored observations are censored at the preceding uncensored time. This estimation technique, however, appears to be computationally cumbersome, especially for large datasets. Another estimation procedure for the Cox model was suggested by Clayton and Cuzick (1985). Their method is a two step algorithm that iterates between estimating the baseline cumulative hazard function using the Breslow's estimator and estimating the model coefficients by maximizing the full log likelihood. This algorithm may not converge, but if it does it converges to the maximum partial likelihood estimate and the Breslow's estimate of the baseline cumulative hazard.

The Cox model also belongs to a class of semiparametric linear transformation models under which an unspecified strictly increasing transformation of the event time is linearly related to the covariates with completely specified error distribution (Doksum, 1987). For the Cox model, the transformation function is the natural logarithm of the baseline cumulative hazard function and the error follows a standard extreme value distribution. Procedures for estimating linear transformation models with right censored data have been developed by several authors. Cheng et al. (1995) proposed a general estimating function for the regression coefficients under the assumption that censoring is independent of covariates. This method was further studied and modified by Cheng et al. (1997) and Fine et al. (1998). Later, Chen et al. (2002) discarded the independent censoring assumption and derived martingale type estimating equations for both the transformation function and model parameters. The resulting coefficient estimator reduces to the Cox's maximum partial likelihood estimator in the special case of the Cox model.

In this paper, we introduce a new estimation procedure for the Cox proportional hazards model. The method uses the covariance matrix of model covariates and alternates between estimating the baseline cumulative hazard function and estimating model parameters. We demonstrate that the estimating equation for model parameters has a form similar to the least squares estimate in the linear regression model. We focus mainly on time independent covariates and uncensored data, but we also propose an approach for dealing with right censored observations. The performance of the new estimation procedure is assessed via simulation and also by its application to real data.

Because of the similarity of our proposed method and the least squares method we can provide an alternative representation for the error between parameter estimates in two nested Cox models in terms of a sample covariance matrix of covariates. In general, we arrive at a simple closed form expression depending on the coefficient estimates of the nonoverlapping covariates, sample covariances between overlapping covariates and between overlapping and nonoverlapping covariates, and on sample covariances of overlapping covariates with the two estimated log transformed baseline cumulative hazard functions. Our representation of the specification error in terms of covariances seems to be new.

The rest of the paper is organized as follows. Section 2 recapitulates the basic facts about the Cox proportional hazards model and introduces the notation. In Section 3 the new estimation procedure for the Cox model is discussed and in Section 4 the alternative representation for the error between parameter estimates in two nested models is derived. In Section 5 the standard and the new approach for Cox regression are compared, and several numerical examples on their performance are given. The application of the proposed method to a real dataset is illustrated in Section 6. Finally, Section 7 contains some concluding remarks and future work.

2. The Cox proportional hazards model

Let T be a positive continuous random variable representing the time to event of interest and $X^{\mathrm{T}} = (X_1, \ldots, X_k)$ be a vector of time independent covariates. Throughout the paper, the superscript T stands for the transpose. The proportional hazards model (Cox, 1972) assumes that the hazard function of T given $X = x$ takes the form:

$$\lambda(t|x) = \lambda_0(t) \exp\left(\beta^{\mathrm{T}} x\right), \tag{1}$$

where $\lambda_0(t)$ is an arbitrary and unspecified baseline hazard function and $\beta^{\mathrm{T}} = (\beta_1, \ldots, \beta_k)$ is a vector of unknown regression coefficients. The associated conditional survival function is

$$S(t|x) = \exp\left\{-\Lambda_0(t) \exp\left(\beta^{\mathrm{T}} x\right)\right\},$$

where $\Lambda_0(t) = \int_0^t \lambda_0(u) du$ is so called baseline cumulative hazard function.

When T is subject to right censoring, the observations consist of independent identically distributed triples $(y_i, \delta_i, x_i^{\mathrm{T}})$ $(i = 1, \ldots, n)$, where $y_i = min(t_i, c_i)$, $\delta_i = I(t_i \leq c_i)$, t_i, c_i and $x_i^{\mathrm{T}} = (x_{1i}, \ldots, x_{ki})$ are respectively the event time, censoring time and the vector of covariate values for ith individual, and $I(\cdot)$ is the indicator function. Cox (1972, 1975) suggested to estimate β by maximizing the partial likelihood function

$$PL(\beta) = \prod_{i=1}^{n} \left\{ \frac{\exp\left(\beta^{\mathrm{T}} x_i\right)}{\sum_{j \in R_i} \exp\left(\beta^{\mathrm{T}} x_j\right)} \right\}^{\delta_i},$$

where $R_i = \{j : y_j \geq y_i\}$ is the risk set at time y_i $(i = 1, \ldots, n)$. The maximum partial likelihood estimator of β is consistent and asymptotically normal (Tsiatis, 1981; Andersen and Gill, 1982).

Estimation of β is typically followed by the estimation of the baseline cumulative hazard function. The estimator of $\Lambda_0(t)$ proposed by Breslow (1972, 1974) is given by

$$\hat{\Lambda}_0^{BR}(t, \hat{\beta}) = \sum_{i: y_i \leq t} \frac{\delta_i}{\sum_{j \in R_i} \exp\left(\hat{\beta}^{\mathrm{T}} x_j\right)}, \tag{2}$$

where $\hat{\beta}$ is the estimator for β in (1). The asymptotic properties of this estimator were established by Andersen and Gill (1982). In general, the estimates of β and $\Lambda_0(t)$ are used to obtain the summary statistics of the event time distribution for an individual with particular values of covariates.

3. Least squares type estimation for the Cox model

Throughout this section we will assume that $\log \Lambda_0(T)$ has a finite second moment and X is a nondegenerate random vector in $\mathbb{R}^k$ with a density function $f_X(x)$ and a covariance matrix C. Here, the log stands for the natural logarithm.

Define $T_x = (T|X = x)$, which is equivalent to saying that T_x is a random variable such that $P(T_x \leq t) = P(T \leq t|X = x)$. Then, it follows from (1) that the cumulative distribution function of T_x is $1 - S(t|x)$. It can be easily shown that with X fixed at x, the survival function of T_x, considered as a function of the random variable T_x, is a uniform random variable on the interval (0,1). This implies that under the Cox model (1) $S(T_x|x) = \exp\{-\Lambda_0(T_x)\exp(\beta^{\mathrm{T}}x)\}$ is uniformly distributed on (0,1), leading to the following definition.

Definition 1.
Let (T, X^{T}) be a random vector such that T is a positive continuous random variable and X is a nondegenerate random vector in $\mathbb{R}^k$. Moreover, let $T_x = (T|X = x)$, $\beta \in \mathbb{R}^k$ and let $\Lambda_0 : \mathbb{R}_+ \to \mathbb{R}_+$ be a strictly increasing continuous function with the property $\lim_{s\to\infty} \Lambda_0(s) = \infty$. We say that (T, X^{T}) satisfies a Cox model with a parameter vector β and the baseline cumulative hazard function Λ_0, and write $(T, X^{\mathrm{T}}) \in Cox(\beta, \Lambda_0)$, if for all $x \in \mathbb{R}^k$ the random variable $U_x = \exp\{-\Lambda_0(T_x)\exp(\beta^{\mathrm{T}}x)\}$ is uniformly distributed on (0,1).

In the first theorem, we prove that a random variable U defined as

$$U = \exp\{-\Lambda_0(T)\exp(\beta^{\mathrm{T}}X)\}$$

has also a uniform distribution and is independent of X. Moreover, we show that in the uncensored case there is at most one Cox model for a given set of covariates.

Theorem 1. *Let (T, X^{T}) be a random vector such that:*

1. *T is a positive continuous random variable,*

2. *X is a nondegenerate random vector in $\mathbb{R}^k$,*
3. *there exists $\beta \in \mathbb{R}^k$ and a strictly increasing continuous function $\Lambda_0 : \mathbb{R}_+ \to \mathbb{R}_+$ with the property $\lim_{s\to\infty} \Lambda_0(s) = \infty$ such that $(T, X^{\mathrm{T}}) \in Cox(\beta, \Lambda_0)$.*

Then:

(i) *random variable $U = \exp\{-\Lambda_0(T)\exp(\beta^{\mathrm{T}}X)\}$ follows a uniform distribution on (0,1) and is independent of X,*

(ii) *if $(T, X^{\mathrm{T}}) \in Cox(\beta^*, \Lambda_0^*)$ with $\beta^* \in \mathbb{R}^k$ and a strictly increasing continuous function $\Lambda_0^* : \mathbb{R}_+ \to \mathbb{R}_+$ such that $\lim_{s\to\infty} \Lambda_0^*(s) = \infty$, then $\beta^* = \beta$ and $\Lambda_0(s) = \Lambda_0^*(s)$ for all $s > 0$.*

Proof of Theorem 1(i). Since $(T, X^{\mathrm{T}}) \in Cox(\beta, \Lambda_0)$,

$$U_x = \exp\{-\Lambda_0(T_x)\exp(\beta^{\mathrm{T}}x)\}$$

is uniform on (0,1), where $T_x = (T|X = x)$. Thus, the conditional density of U given $X = x$ is 1, i.e. $f_{U|X}(u \mid x) = 1$, and

$$f_U(u) = \int f_{UX}(u,x)dx = \int f_{U|X}(u \mid x)f_X(x)dx = \int f_X(x)dx = 1$$

showing that U is uniformly distributed. Also,

$$f_{UX}(u,x) = f_{U|X}(u \mid x)f_X(x) = f_U(u)f_X(x)$$

showing that U is independent of X.

Proof of Theorem 1(ii). By assumptions 3 and (ii),

$$\exp\{-\Lambda_0(T_x)\exp(\beta^{\mathrm{T}}x)\} \sim \exp\{-\Lambda_0^*(T_x)\exp(\beta^{*\mathrm{T}}x)\} \text{ for all } x \in \mathbb{R}^k,$$

where "$\sim$" means "has the same distribution as". Hence,

$$\Lambda_0(T_x) \sim \Lambda_0^*(T_x)\exp\{(\beta^* - \beta)^{\mathrm{T}}x\} \text{ for all } x \in \mathbb{R}^k$$

or, since Λ_0 and Λ_0^* are stictly increasing,

$$P\left(T_x \le \Lambda_0^{-1}(r)\right) = P\left(T_x \le \Lambda_0^{*-1}\left[r\exp\left\{(\beta - \beta^*)^{\mathrm{T}} x\right\}\right]\right) \text{ for all r} > 0, x \in \mathbb{R}^k.$$

Since T_x is continuous, the above equality can hold if and only if

$$r = \Lambda_0\left(\Lambda_0^{*-1}\left[r\exp\left\{(\beta - \beta^*)^{\mathrm{T}} x\right\}\right]\right) \text{ for all r} > 0, x \in \mathbb{R}^k.$$

This can hold for all x only if $\beta = \beta^*$. In that case, $r = \Lambda_0\left(\Lambda_0^{*-1}(r)\right)$ for all r > 0. Since Λ_0 and Λ_0^* are strictly increasing and continuous, this can hold only if $\Lambda_0(s) = \Lambda_0^*(s)$ for all $s > 0$.

The independence of U and X is crucial for the development of the new estimation procedure. It allows establishing the relation between $\Lambda_0(T)$ and β which forms the basis for this procedure. Theorem 2 gives this relation.

Theorem 2.
Let (T, X^{T}) be a random vector, where T is a positive continuous random variable and X is a nondegenerate random vector in $\mathbb{R}^k$ with covariance matrix C. Moreover, let $(T, X^{\mathrm{T}}) \in Cox(\beta, \Lambda_0)$ with $\beta \in \mathbb{R}^k$ and a strictly increasing continuous function $\Lambda_0 : \mathbb{R}_+ \to \mathbb{R}_+$ such that $\lim_{s\to\infty} \Lambda_0(s) = \infty$ and $\log \Lambda_0(T)$ has a finite second moment. Then

$$cov(\log \Lambda_0(T), X_j) = -cov\,(\beta^{\mathrm{T}} X, X_j) \quad (j = 1, \ldots, k) \tag{3}$$

and

$$\beta = -C^{-1} L, \tag{4}$$

where cov denotes covariance, $L^{\mathrm{T}} = (L_1, \ldots, L_k)$ and $L_j = cov(\log \Lambda_0(T), X_j)$ $(j = 1, \ldots, k)$.

Proof of Theorem 2. To show (3), Theorem 1 yields

$$\log(-\log U) = \log \Lambda_0(T) + \beta^{\mathrm{T}} X,$$

where U is uniform on (0,1) and independent of X. Thus,

$$cov\,(\log \Lambda_0(T), X_j) + cov\,(\beta^{\mathrm{T}} X, X_j) = cov\,(\log(-\log U), X_j) = 0$$

and

$$cov\,(\log \Lambda_0(T), X_j) = -cov\,(\beta^{\mathrm{T}} X, X_j) \;\; (j = 1, \ldots, k).$$

To show (4) define L_j as $cov(\log \Lambda_0(T), X_j)$ $(j = 1, \ldots, k)$. Putting $L^{\mathrm{T}} = (L_1, \ldots, L_k)$ the last equation can be rewritten as $L = -C\beta$. Therefore, since C^{-1} exists, $\beta = -C^{-1}L$.

We now demonstrate how relation (4) can be used to estimate regression coefficients in the Cox model (1). Uncensored and censored data are discussed separately.

Equation (3) implies restrictions on the joint probability distribution of (T, X^{T}) in the population, whereas (4) gives parameter vector that satisfies these restrictions. Hence, if the event time T were fully observable, and the probability distribution of (T, X^{T}) as well as the functional representation of

$\Lambda_0(\cdot)$ were known, β could be found directly using (4). In practice, however, the distribution of (T, X^{T}) is only known through a random sample (t_i, x_i^{T}) $(i = 1, \ldots, n)$. In this case, the most intuitive estimator for β is the method of moments estimator obtained by replacing the unknown population moments in (4) with their sample counterparts as follows:

$$\hat{\beta}_n = -C_n^{-1} L_n, \tag{5}$$

where C_n is the sample covariance matrix of X, L_n consists of sample covariances $cov_n(\log \Lambda_0(T), X_j)$ $(j = 1, \ldots, k)$ and $\Lambda_0(T)$ is a random variable whose sample realizations are $\Lambda_0(t_i)$ $(i = 1, \ldots, n)$. Note that there is no difference if we use n or $n - 1$ as a denominator in the sample covariance formula.

The estimating equation (5) has a familiar form of a least squares estimator in the linear regression model. Indeed, define the dependent variable Y^* and the predictor variables $X_1^*, \ldots, X_k^*$ as follows:

$$\begin{aligned} y_i^* &= \tfrac{1}{\sqrt{n-1}} \left(\log \Lambda_0(t_i) - \overline{\log \Lambda_0(T)} \right), \\ x_{ji}^* &= \tfrac{1}{\sqrt{n-1}} \left(x_{ji} - \overline{X}_j \right) \quad (j = 1, \ldots, k; i = 1, \ldots, n), \end{aligned} \tag{6}$$

where $\overline{\log \Lambda_0(T)}$ and $\overline{X}_j$ are the sample means of $\log \Lambda_0(T)$ and X_j observations, respectively. It is easy to verify that the matrix of sums of squares and cross products of the transformed covariates is C_n and that the sum of cross products between X_j^* and Y^* is $cov_n(\log \Lambda_0(T), X_j)$ $(j = 1, \ldots, k)$. Consequently, the least squares estimator for parameters $\beta_1^*, \ldots, \beta_k^*$ in the linear regression model

$$y_i^* = -\beta_1^* x_{1i}^* - \ldots - \beta_k^* x_{ki}^* + \epsilon_i^* \ (i = 1, \ldots, n),$$

where ϵ_i^* is an error term, is given as

$$\hat{\beta}_n^* = -C_n^{-1} L_n,$$

which is equivalent to (5). The reason why there is no intercept in this model is that if it were included its estimate would always be zero. Thus, provided that the baseline cumulative hazard function is known, solving for β using (5) is equivalent to fitting a no intercept linear model to survival data transformed according to (6) and multiplying the coefficients obtained by -1. As a result, the solution for β when the representation of $\Lambda_0(\cdot)$ is known will have a closed form and will be unique.

In most cases, however, the baseline cumulative hazard function is unspecified. By analogy to the classical survival analysis involving the Cox model, we propose to estimate $\Lambda_0(t)$ based on data using the Breslow estimator presented in Section 2. Since this estimator depends on the unknown parameter vector, plugging it in place of $\Lambda_0(t)$ in formula (5) leads to an iterative scheme for estimating β. The main steps of this iterative procedure can be summarized as follows:

Step 1. Obtain the initial estimate $\beta^{(1)}$ of $\hat{\beta}_n$.
Step 2. For $s \geq 1$ estimate $\Lambda_0(t)$ based on data and $\beta^{(s)}$ using (2).
Step 3. Calculate entries of $L_n^{\beta^{(s)}}$ as $cov_n(\log \hat{\Lambda}_0^{BR}\left(T, \beta^{(s)}\right), X_j)$ $(j = 1, \ldots, k)$.
Step 4. Given $L_n^{\beta^{(s)}}$ compute $\beta^{(s+1)}$ as $\beta^{(s+1)} = -C_n^{-1} L_n^{\beta^{(s)}}$.
Step 5. Repeat steps 2 - 4 until a convergence criterion is satisfied.

Our numerical experiments show that this iterative procedure converges and it always leads to the same solution, regardless of the initial estimate $\beta^{(1)}$. Only the number of iterations needed to reach this solution differs. Unfortunately, we presently have no proof of this result. We also observe that the coefficient estimates in the correctly specified model obtained using the iterative procedure appear consistent with their true values and do not differ much from the maximum partial likelihood estimates, both in terms of the mean parameter values and standard errors. Moreover, if applied to estimate a Cox model whose covariates are a subset of those used to generate the data, then the iterative method may produce smaller bias than the maximum partial likelihood method. However, this holds only if the studied covariates are independent of each other and of excluded covariates.

Since formula (4) results from the association between the transformed event time T and covariates, the iterative estimation method just described is suitable only for the uncensored data, that is, when the event time for every individual in the study is known. Therefore, the only way to make the new method work with right censored data is to approximate the unobserved event times by reasonable values and to treat these approximate values as actual event times in the calculations.

Because the censoring time itself is likely to underestimate the true event time, we propose to impute the event time for each censored person by its conditional expectation given the observed data. That is, if t_i is the true event time for an individual with covariate values x_i censored by c_i, then the censored value c_i will be replaced by the conditional expectation $E(t_i|t_i > c_i, x_i)$, which equals to $c_i + \int_{c_i}^{\infty} S(u|x_i)du/S(c_i|x_i)$ and can be estimated by

$$c_i + \frac{1}{\hat{S}(c_i|x_i)} \left\{ \hat{S}(c_i|x_i)(y_i^* - c_i) + \sum_{\substack{j: y_j > c_i \\ y_j \neq \text{largest time}}} \hat{S}(y_j|x_i)\delta_j(y_j^* - y_j) \right\}, \qquad (7)$$

where c_i corresponds to y_i with $\delta_i = 0$, $\hat{S}(\cdot|x_i)$ is the estimated survival function for an individual with covariate values x_i and $y_j^* = min\{y \in \{y_1, \dots, y_n\} : y \text{ is uncensored}, y > y_j\}$. Note that the largest observation has to be treated as uncensored for this calculation. This is due to the fact that the survival function is undefined beyond the largest event time. Also, one has to bear in mind that the estimate of $S(y|x_i)$ depends on β. Hence, every time β changes a new imputed dataset is created. Putting this information together, we obtain an iterative procedure for estimating β which proceeds as follows:

Step 1. Redefine the largest observed time as uncensored.
Step 2. Choose the initial estimate $\beta^{(1)}$ of $\hat{\beta}_n$.
Step 3. For $s \geq 1$ impute the original data using (7) and $\beta^{(s)}$. Denote the imputed sample as $(y_i^I, x_i^{\mathrm{T}})$ $(i = 1, \dots, n)$.
Step 4. Estimate $\Lambda_0(t)$ based on $\beta^{(s)}$ and the imputed sample.
Step 5. Calculate $cov_n(\log \hat{\Lambda}_0^{BR}\left(Y^I, \beta^{(s)}\right), X_j)$ $(j = 1, \dots, k)$ being the entries of $L_n^{\beta^{(s)}}$.
Step 6. Compute $\beta^{(s+1)}$ as $\beta^{(s+1)} = -C_n^{-1} L_n^{\beta^{(s)}}$.
Step 7. Repeat steps 3 - 6 until convergence is achieved.

Note that the Breslow estimator (2) is used twice in this procedure. First, it is applied in step 3 to estimate the survival function for each censored individual. Then, it is used again in step 4 to estimate $\Lambda_0(t)$ for the imputed (uncensored) sample.

Simulation studies reveal that the estimation procedure proposed for censored data performs reasonably well in comparison to the maximum partial likelihood method. Although the new estimator is not generally consistent, which is due to the fact that it may produce asymptotically biased parameter estimates, its bias for each sample size is fairly small. However, both the bias and the standard error of the maximum partial likelihood estimator seem to be smaller in most cases than the corresponding bias and standard error of the new estimator.

4. Model specification error

Consider two specifications of the Cox regression model:

$$\lambda(t|x) = \lambda_0(t)\exp(\beta^{\mathrm{T}}x) = \lambda_0(t)\exp(\beta_{(1)}^{\mathrm{T}}x_{(1)} + \beta_{(2)}^{\mathrm{T}}x_{(2)}) \tag{8}$$

and

$$\lambda(t|x_{(1)}) = \lambda_0^*(t)\exp(\beta^{*\mathrm{T}}x_{(1)}), \tag{9}$$

where $\lambda_0(t)$ and $\lambda_0^*(t)$ are the baseline hazard functions, $x^{\mathrm{T}} = [x_{(1)}^{\mathrm{T}}, x_{(2)}^{\mathrm{T}}]$, $x_{(1)}^{\mathrm{T}} = (x_1, \ldots, x_p)$ $(p < k)$ contains the values of covariates common to the two models, $x_{(2)}^{\mathrm{T}} = (x_{p+1}, \ldots, x_k)$ includes values of covariates unique to model (8), $\beta^{\mathrm{T}} = [\beta_{(1)}^{\mathrm{T}}, \beta_{(2)}^{\mathrm{T}}]$, $\beta_{(1)}^{\mathrm{T}} = (\beta_1, \ldots, \beta_p)$, $\beta_{(2)}^{\mathrm{T}} = (\beta_{p+1}, \ldots, \beta_k)$ and $\beta^{*\mathrm{T}} = (\beta_1^*, \ldots, \beta_p^*)$. Here the model (8) is a valid proportional hazards model.

Our interest in this section is to derive the formula for the error between the corresponding parameter estimates of the two models using the iterative estimation procedure introduced in Section 3. We consider both uncensored and censored data, each containing n observations generated from the model (8). Furthermore, we suppress the explicit dependence of the various quantities on n for the clarity of notation.

We start with the uncensored case. Applying the new estimation procedure to models (8) and (9) we obtain the following coefficient estimates:

$$\hat{\beta} = -C^{-1}L^{\hat{\beta}} = \begin{bmatrix} \hat{\beta}_{(1)} \\ \hat{\beta}_{(2)} \end{bmatrix} \tag{10}$$

and

$$\hat{\beta}^* = -C_{11}^{-1}L^{\hat{\beta}^*}, \tag{11}$$

where C is the sample covariance matrix of $X^{\mathrm{T}} = [X_{(1)}^{\mathrm{T}}, X_{(2)}^{\mathrm{T}}]$ partitioned as

$$C = \begin{bmatrix} C_{11} & C_{12} \\ C_{12}^{\mathrm{T}} & C_{22} \end{bmatrix}$$

with $C_{ij} = cov_n\left(X_{(i)}, X_{(j)}\right)$ $(i, j = 1, 2)$, whereas $L^{\hat{\beta}}$ and $L^{\hat{\beta}^*}$ are respectively column vectors of sample covariances $cov_n(\log\hat{\Lambda}_0^{BR}(T, \hat{\beta}), X_i)$ $(i = 1, \ldots, k)$

and $cov_n(\log \hat{\Lambda}_0^{BR}(T, \hat{\beta}^*), X_j)$ $(j = 1, \ldots, p)$. Subtracting (10) from (11) we can derive the explicit formula for the error between the coefficient estimates in models (8) and (9) which reads

$$\hat{\beta}^* - \hat{\beta}_{(1)} = -C_{11}^{-1}\left(L^{\hat{\beta}^*} - L^{\hat{\beta}}_{(1)} - C_{12}\hat{\beta}_{(2)}\right), \tag{12}$$

where $L^{\hat{\beta}}_{(1)}$ is a column vector consisting of the first p rows of $L^{\hat{\beta}}$.

We see that under the new method the error, which is simply a difference between two known vectors $\hat{\beta}^*$ and $\hat{\beta}_{(1)}$, can also be calculated using the right hand side of (12). The alternative error representation, however, is more than just a calculation rule. First, it shows how the error depends on the correlations between the covariates, the coefficient estimates of the nonoverlapping covariates and the estimated baseline cumulative hazard functions. Second, it allows identifying and grouping factors that contribute to the error's magnitude. These factors include: the coefficient estimates of the nonoverlapping covariates, the sample covariances between overlapping covariates and between overlapping and nonoverlapping covariates, and the sample covariances of overlapping covariates with the two estimated log transformed baseline cumulative hazard functions. Finally, it gives insight, not otherwise available, into the effect of the dependencies among covariates on the error associated with selection of model variables. The use of (12) to predict the sign of the error is discussed below.

Suppose first that the covariates $X_{(1)}$ are uncorrelated with each other and with omitted covariates $X_{(2)}$. Then, the error becomes

$$\hat{\beta}^* - \hat{\beta}_{(1)} = -\left[\frac{cov_n(\log\frac{\hat{\Lambda}_0^{BR}(T,\hat{\beta}^*)}{\hat{\Lambda}_0^{BR}(T,\hat{\beta})}, X_1)}{var_n(X_1)}, \ldots, \frac{cov_n(\log\frac{\hat{\Lambda}_0^{BR}(T,\hat{\beta}^*)}{\hat{\Lambda}_0^{BR}(T,\hat{\beta})}, X_p)}{var_n(X_p)}\right]^{\mathrm{T}}, \tag{13}$$

which is a nonzero vector whose entries are calculated from the complete data. In practice, however, this data is unavailable and hence the signs of $cov_n(\log\frac{\hat{\Lambda}_0^{BR}(T,\hat{\beta}^*)}{\hat{\Lambda}_0^{BR}(T,\hat{\beta})}, X_j)$ $(j = 1, \ldots, p)$ and ultimately the signs of the components of (13) cannot be readily determined. Therefore, the only conclusion that can be drawn from (13) is that when the studied and excluded covariates are uncorrelated the coefficient estimates in the misspecified model are burdened with an error. This result is in accord with the previous findings of Gail et al. (1984), Struthers and Kalbfleisch (1986) and Bretagnolle and Huber-Carol (1988) derived using the properties of the partial likelihood function.

An additional result of these authors is that the error caused by omitting independent covariates is always biased towards zero, which, as explained before, cannot be deduced under the new estimation method.

The error formula (12) does not simplify when all covariates in X are correlated. Moreover, as in the independent covariate case, one cannot say anything about the magnitudes and the signs of elements of $L^{\hat{\beta}^*} - L^{\hat{\beta}}_{(1)}$ unless the complete data is available. This makes the prediction of the direction of the error impossible. However, as seen in the examples, the sign of the error under the correlated covariates can be well predicted using an expression that omits the terms involving the baseline cumulative hazard functions, namely $C_{11}^{-1}C_{12}\hat{\beta}_{(2)}$. The heuristic argument for neglecting these terms is given below.

It is apparent that $L^{\hat{\beta}^*}$ and $L^{\hat{\beta}}_{(1)}$ can be omitted from (12) only if the baseline cumulative hazard functions under models (8) and (9) are proportional, that is, if $\Lambda_0^*(t) = \alpha\Lambda_0(t)$ for every $t > 0$, where α is a constant. In general, this proportionality statement is not true, but one can show that it holds approximately under specific circumstances.

Let $g(x_{(2)}|x_{(1)})$ be the conditional density of $X_{(2)}$ given $X_{(1)} = x_{(1)}$. Then, by elementary probability arguments, the conditional hazard function of T given $X_{(1)}$ is

$$\lambda(t|x_{(1)}) = \frac{f(t|x_{(1)})}{S(t|x_{(1)})} = \frac{\int f(t|x_{(1)}, x_{(2)})g(x_{(2)}|x_{(1)})dx_{(2)}}{\int S(t|x_{(1)}, x_{(2)})g(x_{(2)}|x_{(1)})dx_{(2)}}. \tag{14}$$

Substituting $f(t|x_{(1)}, x_{(2)}) = \lambda(t|x_{(1)}, x_{(2)})S(t|x_{(1)}, x_{(2)})$ and

$$S(t|x_{(1)}, x_{(2)}) = \exp\left\{-\Lambda_0(t)\exp(\beta_{(1)}^{\mathrm{T}}x_{(1)} + \beta_{(2)}^{\mathrm{T}}x_{(2)})\right\}$$

into (14) yields

$$\lambda(t|x_{(1)}) = \lambda_0(t)e^{\beta_{(1)}^{\mathrm{T}}x_{(1)}}h(t; x_{(1)}), \tag{15}$$

where

$$h(t; x_{(1)}) = \frac{\int e^{\beta_{(2)}^{\mathrm{T}}x_{(2)}}e^{-\Lambda_0(t)\exp(\beta_{(1)}^{\mathrm{T}}x_{(1)}+\beta_{(2)}^{\mathrm{T}}x_{(2)})}g(x_{(2)}|x_{(1)})dx_{(2)}}{\int e^{-\Lambda_0(t)\exp(\beta_{(1)}^{\mathrm{T}}x_{(1)}+\beta_{(2)}^{\mathrm{T}}x_{(2)})}g(x_{(2)}|x_{(1)})dx_{(2)}}.$$

Hence, the baseline hazard function under model (9) takes the form

$$\lambda_0^*(t) = \lambda(t|X_{(1)} = 0) = \lambda_0(t)\frac{\int z_1(t; x_{(2)})g(x_{(2)}|X_{(1)} = 0)dx_{(2)}}{\int z_2(t; x_{(2)})g(x_{(2)}|X_{(1)} = 0)dx_{(2)}},$$

where

$$z_1(t;x_{(2)}) = e^{\beta_{(2)}^{\mathrm{T}} x_{(2)}} e^{-\Lambda_0(t)\exp(\beta_{(2)}^{\mathrm{T}} x_{(2)})}$$

and

$$z_2(t;x_{(2)}) = e^{-\Lambda_0(t)\exp(\beta_{(2)}^{\mathrm{T}} x_{(2)})}.$$

Equivalently

$$\lambda_0^*(t) = \lambda_0(t) \frac{E\left[z_1(t;X_{(2)})|X_{(1)}=0\right]}{E\left[z_2(t;X_{(2)})|X_{(1)}=0\right]}, \tag{16}$$

where $E\left[\cdot|X_{(1)}=0\right]$ denotes the conditional expectation with respect to the conditional distribution of $X_{(2)}$ given $X_{(1)} = 0$. To show that $\Lambda_0^*(t)$ and $\Lambda_0(t)$ are nearly proportional define $\bar{x}_{(2)}$ as $E\left[X_{(2)}|X_{(1)}=0\right]$. Expanding $z_1(t;X_{(2)})$ and $z_2(t;X_{(2)})$ in a Taylor series around $X_{(2)} = \bar{x}_{(2)}$, and then applying the conditional expectation operator to both sides of the resulting polynomials we get that for every $t > 0$

$$E\left[z_1(t;X_{(2)})|X_{(1)}=0\right] = z_1(t;\bar{x}_{(2)}) + E[R_1|X_{(1)}=0],$$

$$E\left[z_2(t;X_{(2)})|X_{(1)}=0\right] = z_2(t;\bar{x}_{(2)}) + E[R_2|X_{(1)}=0],$$

where R_1 and R_2 are the remainder terms of the respective Taylor expansions. Assuming that $E[R_1|X_{(1)}=0]$ and $E[R_2|X_{(1)}=0]$ are small we obtain

$$\lambda_0^*(t) \approx \lambda_0(t) \frac{z_1(t;\bar{x}_{(2)})}{z_2(t;\bar{x}_{(2)})} = \lambda_0(t) e^{\beta_{(2)}^{\mathrm{T}} \bar{x}_{(2)}}.$$

Consequently, $\Lambda_0^*(t) \approx \Lambda_0(t) e^{\beta_{(2)}^{\mathrm{T}} \bar{x}_{(2)}}$ and $cov(\log \Lambda_0^*(t), X_j) \approx cov(\log \Lambda_0(t), X_j)$ $(j = 1, \ldots, p)$. This finally implies that $L^{\hat{\beta}^*} - L_{(1)}^{\hat{\beta}} \approx 0$,

$$\hat{\beta}^* - \hat{\beta}_{(1)} \approx C_{11}^{-1} C_{12} \hat{\beta}_{(2)} \tag{17}$$

and

$$sgn\left(\hat{\beta}^* - \hat{\beta}_{(1)}\right) = sgn\left(C_{11}^{-1} C_{12} \hat{\beta}_{(2)}\right),$$

where sgn is a signum function.

We illustrate the prediction of the sign of the error using the right hand side of (17) on three simple examples. Let $k = 2$ and $p = 1$. Then, the error in the coefficient for X_1 has the same sign as $c_{12}\hat{\beta}_2/v_1$, where c_{12} is the sample covariance between X_1 and X_2 and v_1 is the sample variance of X_1.

Hence, since v_1 is always positive, the sign of $\hat{\beta}_1^* - \hat{\beta}_1$ depends on the signs of both c_{12} and $\hat{\beta}_2$, which in many instances can be pretty well predetermined. Specifically, the error is positive if the signs of c_{12} and $\hat{\beta}_2$ agree, and negative otherwise. This result agrees with and complements the simulation results obtained by Abrahamowicz et al. (2004) and Gerds and Schumacher (2001) for the maximum partial likelihood estimates. In case where $k = 3$ and $p = 1$, $sgn(\hat{\beta}_1^* - \hat{\beta}_1) = sgn(c_{12}\hat{\beta}_2 + c_{13}\hat{\beta}_3)$ and

$$sgn(\hat{\beta}_1^* - \hat{\beta}_1) = \begin{cases} 1 & \text{if } c_{12}, c_{13}, \hat{\beta}_2, \hat{\beta}_3 > 0, \\ & \quad c_{12}, c_{13}, \hat{\beta}_2, \hat{\beta}_3 < 0, \\ & \quad c_{12}, \hat{\beta}_2 > 0, c_{13}, \hat{\beta}_3 < 0, \\ & \quad c_{12}, \hat{\beta}_2 < 0, c_{13}, \hat{\beta}_3 > 0, \\ -1 & \text{if } c_{12}, c_{13} > 0, \hat{\beta}_2, \hat{\beta}_3 < 0, \\ & \quad c_{12}, c_{13} < 0, \hat{\beta}_2, \hat{\beta}_3 > 0, \\ & \quad c_{12}, \hat{\beta}_3 > 0, c_{13}, \hat{\beta}_2 < 0, \\ & \quad c_{12}, \hat{\beta}_3 < 0, c_{13}, \hat{\beta}_2 > 0. \end{cases}$$

As far as we know, this type of result has not appeared in the published literature. Finally, assume that $k = 3$ and $p = 2$. Here, $sgn(\hat{\beta}_1^* - \hat{\beta}_1) = -sgn\{\hat{\beta}_3(c_{12}c_{23} - v_2c_{13})\}$ and $sgn(\hat{\beta}_2^* - \hat{\beta}_2) = -sgn\{\hat{\beta}_3(c_{12}c_{13} - v_1c_{23})\}$. Clearly, the signs of $c_{12}c_{23} - v_2c_{13}$ and $c_{12}c_{13} - v_1c_{23}$ depend on the signs and magnitudes of c_{12}, c_{23}, c_{13}, v_1 and v_2, but once they are known the direction of both errors can be predicted following the reasoning in the first example.

The error formula (12) holds true also for the right censored data. In this case, however, $L_{(1)}^{\hat{\beta}}$ and $L^{\hat{\beta}^*}$ consist of $cov_n(\log\hat{\Lambda}_0^{BR}(Y^I, \hat{\beta}), X_j)$ and $cov_n(\log\hat{\Lambda}_0^{BR}(Y^I, \hat{\beta}^*), X_j)$ $(j = 1, \ldots, p)$ computed from two imputed datasets. In general, the imputed event times and the sort order of rows will differ between these data implying that the above deliberation on the specification error does not carry over to the censored case. Nevertheless, as seen in simulations, the misspecification of the model has the same effect on the parameter estimates whether or not the data is censored. We therefore believe that our conclusions about the direction of the error obtained for the uncensored data case are applicable to the censored case as well.

5. Simulation study

In this section we use numerical experiments to evaluate the performance of the iterative estimation procedure for the Cox model presented in Section 3. Our main goal is to investigate and compare the properties of the new estimator and the maximum partial likelihood estimator under various model specifications with both correlated and uncorrelated covariates. We consider both the uncensored and right censored data cases.

To conduct our simulation study we generate event times from a proportional hazards model with three covariates X_1, X_2 and X_3, the conditional hazard function $\lambda(t|x) = \exp(\beta^{\mathrm{T}}x)$ and $\beta^{\mathrm{T}} = \{(1, 1, -1), (1, -1, 2), (2, 1, 1)\}$. Censoring times are generated from an exponential distribution that gives expected proportions of censored observations of 20% and 50%. The marginal distributions of covariates are all uniform on the interval $[-1, 1]$, while the (conditional) Spearman's rank correlations r used to generate correlated random samples of (X_1, X_2, X_3) are $r(X_1, X_2) = 0.9$, $r(X_2, X_3) = -0.9$ and $r(X_1, X_3|X_2) = -0.9$. The corresponding covariances are $cov(X_1, X_2) = 0.3$, $cov(X_1, X_3) = -0.33$ and $cov(X_2, X_3) = -0.3$. The dependence between the variables is modeled by D-vine with the normal copula assigned to its edges (Kurowicka and Cooke, 2006). For every β we generate 1000 datasets of size $n = 4000, 1000, 100$ and 50.

We apply the maximum partial likelihood method and the new procedure to each simulated dataset and we estimate the regression coefficients in the model with all three covariates (*Model 1*), in the model containing X_1 and X_2 (*Model 2*) and in the model including only X_1 (*Model 3*). For each parameter we calculate the sample mean and standard deviation of its resulting 1000 sample specific estimates. All simulations and calculations are carried out in MATLAB. The source codes are available from authors upon request. The stopping criterion for the iterative procedure is defined through the L_2-norm. We stop iterations when the L_2-norm of a vector of differences between successive estimates of β is less than 10^{-7}. The results for the uncensored cases are described first, followed by results for the censored cases.

Tables 1–3 show sample means and standard deviations of parameter estimates in *Model 1*, *Model 2* and *Model 3* when there is no censoring and covariates are uncorrelated. Each of these tables corresponds to a specific vector of the true coefficient values and each row in each table presents the results for different sample size n. Estimates for β_1, β_2 and β_3 for a given model, method and n are given as rows.

Table 1: Means and standard deviations (numbers in parentheses) of parameter estimates in *Models 1*, *2* and *3* obtained using the maximum partial likelihood method (MPL) and the iterative procedure (ITP), when $\beta^{\mathrm{T}} = (1, 1, -1)$, covariates are uncorrelated and the censoring rate is 0%.

	Model 1		***Model 2***		***Model 3***	
	MPL	*ITP*	*MPL*	*ITP*	*MPL*	*ITP*
	1.00(0.03)	1.00(0.04)	0.83(0.03)	0.88(0.04)	0.71(0.03)	0.80(0.04)
$n = 4000$	1.00(0.03)	1.00(0.04)	0.82(0.03)	0.88(0.04)		
	−1.00(0.03)	−1.00(0.04)				
	1.00(0.06)	1.00(0.08)	0.83(0.06)	0.88(0.07)	0.71(0.06)	0.80(0.07)
$n = 1000$	1.00(0.06)	1.00(0.08)	0.83(0.06)	0.88(0.07)		
	−1.00(0.06)	−1.00(0.08)				
	1.03(0.21)	1.00(0.24)	0.85(0.20)	0.87(0.23)	0.73(0.19)	0.79(0.23)
$n = 100$	1.03(0.21)	1.00(0.24)	0.85(0.20)	0.88(0.22)		
	−1.04(0.21)	−1.01(0.24)				
	1.05(0.32)	1.00(0.36)	0.87(0.31)	0.88(0.33)	0.74(0.29)	0.78(0.32)
$n = 50$	1.07(0.32)	1.03(0.36)	0.89(0.31)	0.90(0.33)		
	−1.06(0.34)	−1.03(0.37)				

Table 2: Means and standard deviations (numbers in parentheses) of parameter estimates in *Models 1*, *2* and *3* obtained using the maximum partial likelihood method (MPL) and the iterative procedure (ITP), when $\beta^{\mathrm{T}} = (1, -1, 2)$, covariates are uncorrelated and the censoring rate is 0%.

	Model 1		***Model 2***		***Model 3***	
	MPL	*ITP*	*MPL*	*ITP*	*MPL*	*ITP*
	1.00(0.03)	1.00(0.04)	0.61(0.03)	0.70(0.04)	0.56(0.03)	0.67(0.04)
$n = 4000$	−1.00(0.03)	−1.00(0.04)	−0.61(0.03)	−0.70(0.03)		
	2.00(0.04)	2.00(0.05)				
	1.00(0.06)	1.00(0.08)	0.62(0.06)	0.70(0.07)	0.56(0.06)	0.67(0.07)
$n = 1000$	−1.00(0.06)	−1.00(0.08)	−0.61(0.06)	−0.70(0.07)		
	2.01(0.08)	2.00(0.10)				
	1.02(0.22)	1.00(0.26)	0.62(0.20)	0.69(0.23)	0.57(0.20)	0.66(0.23)
$n = 100$	−1.03(0.22)	−1.01(0.25)	−0.63(0.19)	−0.70(0.21)		
	2.05(0.26)	2.00(0.31)				
	1.05(0.32)	1.01(0.37)	0.64(0.30)	0.69(0.32)	0.58(0.28)	0.65(0.31)
$n = 50$	−1.04(0.33)	−1.01(0.38)	−0.63(0.30)	−0.68(0.32)		
	2.10(0.40)	2.04(0.46)				

In the case involving all three covariates both estimators perform quite well and similarly. They appear consistent and provide mean parameter values that are close to each other and to the true coefficient values even for small sample sizes. However, it seems that the bias of an estimator resulting from the iterative procedure under *Model 1* is in most cases smaller than the bias of the maximum partial likelihood estimator. On the other hand, the estimated standard deviation of the maximum partial likelihood estimator is

Table 3: Means and standard deviations (numbers in parentheses) of parameter estimates in *Models 1*, *2* and *3* obtained using the maximum partial likelihood method (MPL) and the iterative procedure (ITP), when $\beta^{\mathrm{T}} = (2, 1, 1)$, covariates are uncorrelated and the censoring rate is 0%.

	Model 1		***Model 2***		***Model 3***	
	MPL	*ITP*	*MPL*	*ITP*	*MPL*	*ITP*
	2.00(0.04)	2.00(0.05)	1.67(0.04)	1.76(0.04)	1.46(0.03)	1.60(0.04)
$n = 4000$	1.00(0.03)	1.00(0.04)	0.84(0.03)	0.88(0.04)		
	1.00(0.03)	1.00(0.04)				
	2.00(0.07)	2.00(0.09)	1.67(0.07)	1.75(0.08)	1.46(0.07)	1.60(0.08)
$n = 1000$	1.00(0.06)	0.99(0.08)	0.84(0.06)	0.88(0.07)		
	1.00(0.06)	0.99(0.08)				
	2.04(0.27)	2.01(0.32)	1.70(0.25)	1.75(0.27)	1.48(0.23)	1.59(0.25)
$n = 100$	1.03(0.21)	1.01(0.25)	0.86(0.21)	0.88(0.23)		
	1.03(0.21)	1.00(0.25)				
	2.10(0.40)	2.03(0.46)	1.75(0.37)	1.76(0.39)	1.52(0.34)	1.58(0.35)
$n = 50$	1.06(0.32)	1.02(0.38)	0.88(0.32)	0.88(0.35)		
	1.06(0.32)	1.03(0.37)				

always slightly smaller than the standard deviation of the other estimator. Nevertheless, the difference between estimators is not great and becomes smaller as the the sample size gets larger.

When the covariate X_3 is omitted from *Model 1*, the estimators of β_1 and β_2 become biased towards zero. This holds for any vector of true coefficient values, sample size n and estimation method. The same effect is observed for β_1 when also covariate X_2 (in addition to X_3) is omitted. Generally, the more excluded covariates the larger the bias. Additionally, the exclusion of X_3 and also X_2 does not affect the signs of the coefficients in both submodels of *Model 1*. They are consistent with the signs of their corresponding true values. These observations confirm the theoretical results of Bretagnolle and Huber-Carol (1988), but they do not follow from the error formula derived in Section 4.

As shown in tables, the mean parameter estimates for *Model 2* and *Model 3* obtained with the maximum partial likelihood method and the iterative procedure do not agree as much as for the correct *Model 1*. However, the bias of estimators of coefficients in both misspecified models provided by the iterative procedure is always smaller than the corresponding bias based on the maximum partial likelihood method. Again, as in the correct case, the maximum partial likelihood estimator has a smaller variation.

Not many of the above conclusions hold when the model covariates are correlated. The results of simulations involving uncensored data with correlated covariates are presented in Tables 4–6.

Table 4: Means and standard deviations (numbers in parentheses) of parameter estimates in *Models 1*, *2* and *3* obtained using the maximum partial likelihood method (MPL) and the iterative procedure (ITP), when $\beta^{\mathrm{T}} = (1, 1, -1)$, covariates are correlated and the censoring rate is 0%.

	Model 1		***Model 2***		***Model 3***	
	MPL	*ITP*	*MPL*	*ITP*	*MPL*	*ITP*
	1.00(0.15)	1.00(0.19)	1.89(0.07)	1.90(0.09)	2.72(0.05)	2.76(0.06)
$n = 4000$	1.00(0.06)	1.00(0.08)	1.08(0.06)	1.08(0.08)		
	−1.00(0.15)	−1.00(0.19)				
	1.01(0.30)	1.01(0.38)	1.90(0.14)	1.89(0.17)	2.73(0.09)	2.76(0.11)
$n = 1000$	1.00(0.13)	1.00(0.17)	1.08(0.13)	1.08(0.16)		
	−1.00(0.30)	−0.99(0.38)				
	1.03(1.07)	0.98(1.25)	1.93(0.47)	1.87(0.56)	2.76(0.32)	2.72(0.36)
$n = 100$	1.02(0.45)	1.00(0.54)	1.09(0.44)	1.07(0.53)		
	−1.03(1.07)	−1.02(1.27)				
	1.03(1.58)	1.02(1.80)	1.96(0.68)	1.92(0.83)	2.81(0.47)	2.76(0.54)
$n = 50$	1.07(0.67)	1.02(0.82)	1.14(0.64)	1.09(0.79)		
	−1.07(1.58)	−1.02(1.79)				

Table 5: Means and standard deviations (numbers in parentheses) of parameter estimates in *Models 1*, *2* and *3* obtained using the maximum partial likelihood method (MPL) and the iterative procedure (ITP), when $\beta^{\mathrm{T}} = (1, -1, 2)$, covariates are correlated and the censoring rate is 0%.

	Model 1		***Model 2***		***Model 3***	
	MPL	*ITP*	*MPL*	*ITP*	*MPL*	*ITP*
	1.00(0.15)	1.01(0.19)	−0.79(0.06)	−0.79(0.08)	−1.70(0.04)	−1.75(0.04)
$n = 4000$	−1.00(0.07)	−1.00(0.09)	−1.13(0.07)	−1.14(0.08)		
	2.00(0.15)	2.01(0.19)				
	1.00(0.31)	1.00(0.39)	−0.79(0.13)	−0.80(0.16)	−1.71(0.07)	−1.75(0.08)
$n = 1000$	−1.00(0.13)	−0.99(0.16)	−1.13(0.13)	−1.13(0.16)		
	2.01(0.31)	2.00(0.39)				
	1.06(1.05)	1.04(1.26)	−0.80(0.43)	−0.78(0.51)	−1.72(0.24)	−1.73(0.27)
$n = 100$	−1.04(0.45)	−1.01(0.53)	−1.15(0.44)	−1.14(0.51)		
	2.07(1.06)	2.02(1.28)				
	1.14(1.65)	1.15(1.87)	−0.80(0.64)	−0.74(0.73)	−1.74(0.35)	−1.73(0.38)
$n = 50$	−1.06(0.69)	−1.05(0.79)	−1.19(0.64)	−1.19(0.75)		
	2.18(1.67)	2.12(1.90)				

The mean values of parameter estimates in *Model 1* estimated using the maximum partial likelihood method and the iterative procedure remain close to each other in most cases, but their distance from the true coefficient values increases with decreasing sample size. Furthermore, the biases and estimated standard deviations of both estimators are generally smaller in the case of uncorrelated covariates, reflecting the fact that correlation reduces effective sample size. Nevertheless, both estimators seem to be consistent and in comparison with the maximum partial likelihood estimator the estimator

Table 6: Means and standard deviations (numbers in parentheses) of parameter estimates in *Models 1*, *2* and *3* obtained using the maximum partial likelihood method (MPL) and the iterative procedure (ITP), when $\beta^{\mathrm{T}} = (2, 1, 1)$, covariates are correlated and the censoring rate is 0%.

	Model 1		***Model 2***		***Model 3***	
	MPL	*ITP*	*MPL*	*ITP*	*MPL*	*ITP*
	2.00(0.15)	2.00(0.19)	1.08(0.06)	1.09(0.08)	1.83(0.04)	1.86(0.04)
$n = 4000$	1.00(0.06)	1.00(0.08)	0.91(0.06)	0.91(0.08)		
	1.00(0.14)	1.00(0.19)				
	2.01(0.31)	2.01(0.38)	1.09(0.13)	1.09(0.16)	1.83(0.07)	1.85(0.08)
$n = 1000$	1.00(0.13)	0.99(0.17)	0.91(0.13)	0.90(0.16)		
	1.01(0.31)	1.01(0.38)				
	2.03(1.12)	2.00(1.30)	1.10(0.44)	1.08(0.51)	1.86(0.24)	1.85(0.27)
$n = 100$	1.04(0.45)	1.01(0.54)	0.95(0.45)	0.92(0.53)		
	1.01(1.11)	1.00(1.28)				
	2.18(1.75)	2.16(1.91)	1.13(0.67)	1.10(0.76)	1.90(0.37)	1.87(0.41)
$n = 50$	1.09(0.70)	1.03(0.78)	0.97(0.67)	0.93(0.75)		
	1.13(1.71)	1.14(1.89)				

resulting from the iterative procedure has larger standard deviations.

The exclusion of covariate X_3 which is correlated with X_1 and X_2 does not lead to one systematic effect, but it causes three different effects. These effects are clearly visible for every sample size n and estimation method. They include: (1) bias away from zero of the estimator of β_1 and β_2, (2) the change in sign of the estimator of β_1 and bias away from zero of the estimator of β_2, and (3) bias towards zero of the estimator of β_1 and β_2. The direction of the bias of estimator for β_1 remains the same when also X_2 is excluded. The simulation results for both submodels of *Model 1* do not allow for a fair comparison between estimators with respect to bias. However, they clearly indicate that the maximum partial likelihood method yields smaller standard deviations for parameters in *Model 2* and *Model 3* than the iterative procedure.

It is noteworthy that the sign of the bias of parameter estimators in *Model 2* and *Model 3* agrees with the sign of the elements of the right hand side of formula (17) where the sample covariances and parameter estimates of omitted covariates are replaced by their population counterparts. This holds regardless of the sample size and estimation procedure. Interestingly, if only the signs of the true parameter values of the omitted covariates and not their actual values are used, the direction of the bias can be correctly predicted in all the cases except for *Model 3* when $\beta^T = (2, 1, 1)$. For illustration, take $\beta^T = (1, -1, 2)$. According to Section 4 the signs of the errors in the coefficient estimates for X_1 and X_2 in *Model 2* equal respectively $sgn\{-\hat{\beta}_3(c_{12}c_{23} - v_2 c_{13})\}$

and $sgn\{-\hat{\beta}_3(c_{12}c_{13} - v_1c_{23})\}$. Substituting β_3 in place of $\hat{\beta}_3$ and $cov(X_i, X_j)$ in place of c_{ij} $(i, j = 1, 2, 3)$ we get that the bias of the estimator of β_1 and β_2 will have the same sign as $\beta_3\{var(X_2)cov(X_1, X_3)-cov(X_1, X_2)cov(X_2, X_3)\}$ and $\beta_3\{var(X_1)cov(X_2, X_3) - cov(X_1, X_2)cov(X_1, X_3)\}$, respectively. In this example, $\beta_3 > 0$, $var(X_2)cov(X_1, X_3)-cov(X_1, X_2)cov(X_2, X_3) = -0.02$ and $var(X_1)cov(X_2, X_3) - cov(X_1, X_2)cov(X_1, X_3) = -0.001$ implying that the biases of estimators of β_1 and β_2 caused by omitting X_3 are negative. Following the same logic we can show that the sign of bias of the estimator of β_1 in *Model 3* is the same as the sign of $cov(X_1, X_2)\beta_2 + cov(X_1, X_3)\beta_3$, which is negative.

Tables 7–10 provide simulation results obtained for the right censored data. However, since the performance of the iterative and the maximum partial likelihood methods is similar for all three vectors of true parameter values, only the results for $\beta^{\mathrm{T}} = (1, 1, -1)$ are reported.

Table 7: Means and standard deviations (numbers in parentheses) of parameter estimates in *Models 1, 2* and *3* obtained using the maximum partial likelihood method (MPL) and the iterative procedure (ITP), when $\beta^{\mathrm{T}} = (1, 1, -1)$, covariates are uncorrelated and the expected censoring rate is 20%.

	Model 1		***Model 2***		***Model 3***	
	MPL	*ITP*	*MPL*	*ITP*	*MPL*	*ITP*
	1.00(0.03)	1.04(0.05)	0.82(0.03)	0.91(0.04)	0.71(0.03)	0.83(0.04)
$n = 4000$	1.00(0.03)	1.04(0.04)	0.82(0.03)	0.91(0.04)		
	−1.00(0.03)	−1.04(0.04)				
	1.00(0.07)	1.04(0.09)	0.83(0.07)	0.91(0.09)	0.72(0.07)	0.83(0.09)
$n = 1000$	1.00(0.07)	1.04(0.09)	0.83(0.07)	0.91(0.08)		
	−1.00(0.07)	−1.04(0.09)				
	1.03(0.24)	1.05(0.31)	0.85(0.23)	0.90(0.28)	0.74(0.23)	0.82(0.28)
$n = 100$	1.03(0.24)	1.04(0.30)	0.85(0.24)	0.90(0.28)		
	−1.03(0.24)	−1.05(0.30)				
	1.07(0.37)	1.07(0.46)	0.88(0.34)	0.91(0.39)	0.75(0.32)	0.81(0.37)
$n = 50$	1.06(0.37)	1.06(0.45)	0.88(0.35)	0.91(0.41)		
	−1.08(0.36)	−1.06(0.45)				

One can observe that, unlike the maximum partial likelihood estimator, the estimator of coefficients in *Model 1* resulting from the iterative procedure loses consistency as the expected proportion of censored observations increases. Moreover, for every n, when the censoring rate gets bigger the mean parameter values obtained using the iterative procedure become larger in absolute value, whereas the means of the other estimator remain close to each other and to their respective true values. This immediately implies that the differences between the mean estimates of the two estimators and

Table 8: Means and standard deviations (numbers in parentheses) of parameter estimates in *Models 1*, *2* and *3* obtained using the maximum partial likelihood method (MPL) and the iterative procedure (ITP), when $\beta^{\mathrm{T}} = (1, 1, -1)$, covariates are uncorrelated and the expected censoring rate is 50%.

	Model 1		***Model 2***		***Model 3***	
	MPL	*ITP*	*MPL*	*ITP*	*MPL*	*ITP*
	1.00(0.04)	1.23(0.07)	0.82(0.04)	1.09(0.07)	0.71(0.04)	1.05(0.06)
$n = 4000$	1.00(0.04)	1.23(0.07)	0.82(0.04)	1.09(0.07)		
	−1.00(0.04)	−1.23(0.07)				
	1.00(0.09)	1.22(0.14)	0.83(0.09)	1.08(0.13)	0.72(0.08)	1.03(0.12)
$n = 1000$	1.00(0.09)	1.23(0.14)	0.83(0.09)	1.09(0.13)		
	−1.00(0.09)	−1.22(0.15)				
	1.03(0.30)	1.20(0.45)	0.85(0.29)	1.04(0.42)	0.73(0.28)	0.97(0.39)
$n = 100$	1.03(0.30)	1.20(0.47)	0.85(0.29)	1.03(0.44)		
	−1.03(0.31)	−1.20(0.46)				
	1.08(0.48)	1.21(0.78)	0.90(0.45)	1.03(0.62)	0.75(0.42)	0.94(0.56)
$n = 50$	1.10(0.46)	1.24(0.72)	0.91(0.44)	1.05(0.62)		
	−1.07(0.48)	−1.19(0.75)				

Table 9: Means and standard deviations (numbers in parentheses) of parameter estimates in *Models 1*, *2* and *3* obtained using the maximum partial likelihood method (MPL) and the iterative procedure (ITP), when $\beta^{\mathrm{T}} = (1, 1, -1)$, covariates are correlated and the expected censoring rate is 20%.

	Model 1		***Model 2***		***Model 3***	
	MPL	*ITP*	*MPL*	*ITP*	*MPL*	*ITP*
	1.00(0.17)	1.04(0.22)	1.89(0.08)	1.97(0.10)	2.72(0.05)	2.86(0.06)
$n = 4000$	1.00(0.07)	1.05(0.10)	1.08(0.07)	1.13(0.09)		
	−0.99(0.17)	−1.04(0.22)				
	1.00(0.34)	1.05(0.46)	1.90(0.15)	1.97(0.20)	2.73(0.10)	2.86(0.13)
$n = 1000$	1.00(0.15)	1.04(0.19)	1.08(0.14)	1.12(0.19)		
	−1.00(0.34)	−1.03(0.46)				
	0.94(1.20)	0.94(1.53)	1.90(0.53)	1.94(0.69)	2.75(0.34)	2.83(0.44)
$n = 100$	1.05(0.52)	1.05(0.67)	1.13(0.50)	1.14(0.65)		
	−1.09(1.19)	−1.14(1.51)				
	1.06(1.90)	1.06(2.35)	1.91(0.77)	1.93(1.05)	2.77(0.49)	2.83(0.78)
$n = 50$	1.08(0.80)	1.09(1.00)	1.14(0.75)	1.16(0.95)		
	−0.99(1.94)	−1.00(2.37)				

the bias of an estimator based on the iterative procedure increase with increasing censoring rate. The discrepancies, however, are not very large. As also seen in the tables, higher censoring level denotes larger standard errors of both estimators. Nonetheless, as in the uncensored case, the variation of the estimates produced by the maximum partial likelihood method is always smaller.

The aforementioned observations hold true in both correlated and uncorrelated covariate cases and all of them except two regarding the consistency and the biases of the two estimators also apply to the submodels of *Model 1*.

Table 10: Means and standard deviations (numbers in parentheses) of parameter estimates in *Models 1*, *2* and *3* obtained using the maximum partial likelihood method (MPL) and the iterative procedure (ITP), when $\beta^{\mathrm{T}} = (1, 1, -1)$, covariates are correlated and the expected censoring rate is 50%.

	Model 1		***Model 2***		***Model 3***	
	MPL	*ITP*	*MPL*	*ITP*	*MPL*	*ITP*
	1.00(0.21)	1.19(0.33)	1.89(0.10)	2.26(0.15)	2.72(0.07)	3.27(0.10)
$n = 4000$	1.00(0.09)	1.19(0.14)	1.08(0.09)	1.28(0.14)		
	−1.00(0.21)	−1.19(0.32)				
	1.00(0.43)	1.19(0.65)	1.90(0.19)	2.25(0.30)	2.73(0.13)	3.26(0.20)
$n = 1000$	1.01(0.19)	1.19(0.30)	1.08(0.19)	1.28(0.29)		
	−1.01(0.42)	−1.19(0.65)				
	1.07(1.45)	1.30(2.18)	1.94(0.65)	2.24(1.00)	2.76(0.43)	3.19(0.66)
$n = 100$	1.02(0.66)	1.14(0.96)	1.08(0.63)	1.21(0.92)		
	−0.99(1.48)	−1.17(2.19)				
	1.04(2.43)	1.06(3.40)	1.99(1.03)	2.20(1.51)	2.82(0.67)	3.09(1.02)
$n = 50$	1.06(1.03)	1.12(1.48)	1.13(0.97)	1.20(1.39)		
	−1.12(2.49)	−1.25(3.59)				

A final observation is that, regardless of the censoring level, the parameter estimates in *Model 2* and *Model 3* obtained using the iterative procedure are always biased in the same direction (relative to the mean estimates in *Model 1*) as the corresponding maximum partial likelihood estimates.

6. Application

In this section, we illustrate the performance of the new estimation procedure on a real dataset containing information about the survival of 228 lung cancer patients from a clinical study at the Mayo Clinic, of whom 165 have observed death times and 63 are censored (Loprinzi et al., 1994; Therneau and Grambsch, 2000). The covariates studied are: age, sex, physician's estimate of the ECOG (Eastern Cooperative Oncology Group) performance score (*ph.ecog*), physician's assessment of the Karnofsky performance score (*ph.karno*), the Karnofsky performance score as rated by the patient (*pat.karno*), the number of calories consumed at meals (*meal.cal*) and the amount of weight lost or gained in the last six months (*wt.loss*). 60 subjects were excluded from the analysis due to missing covariate values. The final dataset involves 168 observations, among which 47 are censored. As before, we analyze both the censored and uncensored dataset.

Tables 11–12 show the results of fitting various models using the maximum partial likelihood method and the iterative procedure in the uncensored and censored case, respectively. The first row in each table lists all the covariates whereas the columns 3–9 give the estimates of regression coefficients for these variables in seven different models. For comparison purposes, we also included the standard errors of the maximum partial likelihood estimates. *Model 1* uses all 7 covariates and the following models are obtained by successively removing one covariate from *Model 1*. At each step, we eliminate a covariate whose omission causes the strongest error in parameter estimates of the remaining covariates. The dash (–) sign denotes that the variable is not included in the current model.

Table 11: Parameter estimates in *Models 1–7* fitted to the uncensored observations in the lung cancer data using the maximum partial likelihood method (MPL) and the iterative procedure (ITP). The numbers in parentheses are standard errors.

		age	*sex*	*ph.ecog*	*ph.karno*	*pat.karno*	*meal.cal*	*wt.loss*
Model 1	*MPL*	0.00 (0.01)	-0.24 (0.21)	0.24 (0.24)	0.00 (0.01)	-0.01 (0.01)	0.00 (0.00)	-0.01 (0.01)
	ITP	0.01	-0.36	0.26	0.00	-0.01	0.00	-0.01
Model 2	*MPL*	0.00 (0.01)	-0.24 (0.21)	0.28 (0.24)	0.00 (0.01)	-0.01 (0.01)	–	-0.01 (0.01)
	ITP	0.02	-0.29	0.31	0.00	-0.01	–	-0.01
Model 3	*MPL*	0.01 (0.01)	-0.17 (0.20)	0.26 (0.24)	0.01 (0.01)	-0.01 (0.01)	–	–
	ITP	0.02	-0.25	0.30	0.00	-0.01	–	–
Model 4	*MPL*	0.00 (0.01)	-0.17 (0.20)	0.18 (0.16)	–	-0.01 (0.01)	–	–
	ITP	0.02	-0.25	0.24	–	-0.01	–	–
Model 5	*MPL*	–	-0.18 (0.20)	0.19 (0.16)	–	-0.01 (0.01)	–	–
	ITP	–	-0.30	0.28	–	-0.01	–	–
Model 6	*MPL*	–	-0.23 (0.20)	0.30 (0.13)	–	–	–	–
	ITP	–	-0.33	0.39	–	–	–	–
Model 7	*MPL*	–	-0.18 (0.20)	–	–	–	–	–
	ITP	–	-0.26	–	–	–	–	–

Table 12: Parameter estimates in *Models 1–7* fitted to the censored and uncensored observations in the lung cancer data using the maximum partial likelihood method (MPL) and the iterative procedure (ITP). The numbers in parentheses are standard errors.

		age	*sex*	*ph.ecog*	*ph.karno*	*pat.karno*	*meal.cal*	*wt.loss*
Model 1	*MPL*	0.01 (0.01)	-0.55 (0.20)	0.73 (0.22)	0.02 (0.01)	-0.01 (0.01)	0.00 (0.00)	-0.01 (0.01)
	ITP	0.02	-0.68	0.71	0.02	-0.01	0.00	-0.01
Model 2	*MPL*	0.01 (0.01)	-0.53 (0.20)	0.44 (0.17)	–	-0.01 (0.01)	0.00 (0.00)	-0.01 (0.01)
	ITP	0.01	-0.69	0.44	–	-0.01	0.00	-0.01
Model 3	*MPL*	0.01 (0.01)	-0.54 (0.20)	0.54 (0.15)	–	–	0.00 (0.00)	-0.01 (0.01)
	ITP	0.01	-0.71	0.54	–	–	0.00	-0.01
Model 4	*MPL*	0.01 (0.01)	-0.50 (0.20)	0.45 (0.14)	–	–	0.00 (0.00)	–
	ITP	0.02	-0.65	0.52	–	–	0.00	–
Model 5	*MPL*	–	-0.51 (0.20)	0.47 (0.13)	–	–	0.00 (0.00)	–
	ITP	–	-0.71	0.56	–	–	0.00	–
Model 6	*MPL*	–	-0.51 (0.20)	0.48 (0.13)	–	–	–	–
	ITP	–	-0.67	0.60	–	–	–	–
Model 7	*MPL*	–	-0.48 (0.20)	–	–	–	–	–
	ITP	–	-0.63	–	–	–	–	–

The results of this example are in a good agreement with the results from the simulation experiments presented in Section 5. In most cases, the estimates of parameters in different models produced by the iterative procedure vary from their respective maximum partial likelihood estimates, what may be partially explained by the small sample size. The differences, however, are not very significant as the new method's estimates fall well within their corresponding 95% confidence intervals around the maximum partial likelihood estimates. The results also show that in both uncensored and censored case the error in the parameter estimates of two consecutive models is not unidirectional and fluctuates from model to model, which is due to the dependence among the covariates (the correlation matrices of model variables in the uncensored and censored case are given in Tables 13–14). Interestingly, in the majority of cases, the signs of errors in coefficient estimates between any two models resulting from the iterative procedure agree with those obtained using the maximum partial likelihood method.

Table 13: Correlation matrix of covariates calculated from uncensored lung cancer data.

	age	*sex*	*ph.ecog*	*ph.karno*	*pat.karno*	*meal.cal*	*wt.loss*
age	1.00	-0.14	0.22	-0.26	-0.15	-0.24	-0.04
sex	-0.14	1.00	0.10	-0.09	0.04	-0.08	-0.17
ph.ecog	0.22	0.10	1.00	-0.84	-0.50	-0.14	0.07
ph.karno	-0.26	-0.09	-0.84	1.00	0.52	0.09	-0.07
pat.karno	-0.15	0.04	-0.50	0.52	1.00	0.16	-0.09
meal.cal	-0.24	-0.08	-0.14	0.09	0.16	1.00	-0.08
wt.loss	-0.04	-0.17	0.07	-0.07	-0.09	-0.08	1.00

Table 14: Correlation matrix of covariates calculated from censored lung cancer data.

	age	*sex*	*ph.ecog*	*ph.karno*	*pat.karno*	*meal.cal*	*wt.loss*
age	1.00	-0.13	0.31	-0.33	-0.24	-0.24	0.05
sex	-0.13	1.00	-0.01	-0.02	0.07	-0.17	-0.17
ph.ecog	0.31	-0.01	1.00	-0.82	-0.54	-0.11	0.18
ph.karno	-0.33	-0.02	-0.82	1.00	0.53	0.06	-0.13
pat.karno	-0.24	0.07	-0.54	0.53	1.00	0.17	-0.18
meal.cal	-0.24	-0.17	-0.11	0.06	0.17	1.00	-0.11
wt.loss	0.05	-0.17	0.18	-0.13	-0.18	-0.11	1.00

7. Concluding remarks and future work

In this paper we introduced a new procedure for estimating the Cox proportional hazards model. The method is iterative in nature and returns the estimates of both the regression coefficients and the baseline cumulative hazard function. As shown in the examples, the new approach imitates the standard maximum partial likelihood method when estimating both the correct and misspecified models based on uncensored survival data with time invariant (dependent or independent) covariates. Work still needs to be done to improve the properties of the new estimator in case of censored observations. The convergence of the iterative scheme also needs to be proved. The main benefit of the new method is decomposing the error between the coefficient estimates in two nested Cox models into the following components: the coefficient estimates of the nonoverlapping covariates, the sample covariances between overlapping covariates and between overlapping and nonoverlapping covariates, and the sample covariances between the overlapping covariates and the two estimated log transformed baseline cumulative hazard functions. We believe this result might help better understand the influence of the dependencies among covariates on the specification error and bias.

Acknowledgement

This research has been a part of the European Union funded project BENERIS (Benefit-Risk Assessment for Food: an Iterative Value-of-Information Approach, Contract No. Food-CT-2006-022936). We would like to thank Prof. Bo Lindqvist for valuable comments and for suggesting imputation procedure to extend our estimation method to the right censored data.

Abrahamowicz, M., du Berger, R., Krewski, D., Burnett, R., Bartlett, G., Tamblyn, R.M., Leffondre, K., 2004. Bias due to aggregation of individual covariates in the Cox regression model. Am. J. Epidemiol. 160, 696–706.

Andersen, P.K., Gill, R.D., 1982. Cox's regression model for counting processes: a large sample study. Ann. Statist. 10, 1100–1120.

Breslow, N.E., 1972. Contribution to the discussion on paper by D. R. Cox. J. R. Stat. Soc. Ser. B Stat. Methodol. 34, 216–217.

Breslow, N.E., 1974. Covariance analysis of censored survival data. Biometrics 30, 89–99.

Bretagnolle, J., Huber-Carol, C., 1988. Effects of omitting covariates in Cox's model for survival data. Scand. J. Statist. 15, 125–138.

Chen, K., Jin, Z., Ying, Z., 2002. Semiparametric analysis of transformation models with censored data. Biometrika 89, 659–668.

Cheng, S.C., Wei, L.J., Ying, Z., 1995. Analysis of transformation models with censored data. Biometrika 82, 835–845.

Cheng, S.C., Wei, L.J., Ying, Z., 1997. Predicting survival probabilities with semiparametric transformation models. J. Am. Statist. Assoc. 92, 227–235.

Clayton, D., Cuzick, J., 1985. The EM algorithm for Cox's regression model using GLIM. J. R. Stat. Soc. Ser. C Appl. Statist. 34, 148–156.

Cox, D.R., 1972. Models and life-tables (with discussion). J. R. Stat. Soc. Ser. B Stat. Methodol. 34, 187–220.

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

Doksum, K.A., 1987. An extension of partial likelihood methods for proportional hazard models to general transformation models. Ann. Statist. 15, 325–345.

Fine, J.P., Ying, Z., Wei, L.J., 1998. On the linear transformation model for censored data. Biometrika 85, 980–986.

Gail, M.H., Wieand, S., Piantadosi, S., 1984. Biased estimates of treatment effect in randomized experiments with nonlinear regressions and omitted covariates. Biometrika 71, 431–444.

Gerds, T.A., Schumacher, M., 2001. On functional misspecification of covariates in the Cox regression model. Biometrika 88, 572–580.

Kalbfleisch, J.D., Prentice, R.L., 1973. Marginal likelihoods based on Cox's regression and life model. Biometrika 60, 267–278.

Kurowicka, D., Cooke, R.M., 2006. Uncertainty analysis with high dimensional dependence modelling. John Wiley and Sons, West Sussex, England.

Lawless, J.F., 1982. Statistical models and methods for lifetime data. John Wiley and Sons, New York.

Loprinzi, C.L., Goldberg, R.M., Su, J.Q., Mailliard, J.A., Kuross, S.A., Maksymiuk, A.W., Kugler, J.W., Jett, J.R., Ghosh, C., Pfeifle, D.M., Wender, D.B., Burch, P.A., 1994. Placebo-controlled trial of hydrazine sulfate in patients with newly diagnosed non-small-cell lung cancer. J. Clin. Oncol. 12, 1126–1129.

Struthers, C.A., Kalbfleisch, J.D., 1986. Misspecified proportional hazard models. Biometrika 73, 363–369.

Therneau, T.M., Grambsch, P.M., 2000. Modeling survival data - extending the Cox model. Springer-Verlag, New York.

Tsiatis, A.A., 1981. A large sample study of Cox's regression model. Ann. Statist. 9, 93–108.