A general class of time-varying coefficients models for right censored data

C. Legrand[a,*], M. Munda[a], P. Janssen[b], L. Duchateau[c]

[a] *Université catholique de Louvain. Institute of Statistics, Biostatistics and Actuarial Sciences (ISBA). Voie du Roman Pays 20, 1348 Louvain-la-neuve, Belgium.*
[b] *Hasselt University. Center for Statistics. Agoralaan, 3590 Diepenbeek, Belgium.*
[c] *Universiteit Gent. Department of Comparative Physiology and Biometrics. Salisburylaan 133, 9820 Merelbeke, Belgium.*

Abstract

While the Cox proportional hazards model is probably the most popular one to model survival data, there exist several alternatives, such as the additive risk model and the proportional odds model. All these models can be studied in an unified framework using a link function $\phi : [0, 1] \to \mathbb{R}$ to model the impact of the covariates on the transformed conditional survival function $\phi(S(y|x))$. In this paper we extend these models to time-dependent coefficients. For various choices of the link function, we discuss possible choices for the time-dependency of the coefficients. This will reveal concrete families of conditional survival functions and result in an easy way to generate data from models with time-dependent regression coefficients. We also discuss the practical implementation of the least squares estimation method proposed by Teodorescu et al. (2010); and we propose a pragmatic approach for the choice of the bandwidth needed to estimate $S(y|x)$ nonparametrically. Results on a

*Corresponding author
Email address: catherine.legrand@uclouvain.be (C. Legrand)
URL: http://www.uclouvain.be/isba (C. Legrand)

real bladder cancer database are used to illustrate how this procedure works in practice and simulations in the context of oncology clinical trials demonstrate usefulness of the procedure in detecting time-dependency in the effect of a continuous covariate.
Keywords: Survival, time-varying coefficients, transformation model, bootstrap, bladder cancer

1. Introduction

Right-censored time-to-event data are often encountered in the field of medicine, especially when working with oncology data. In this context, the time-to-event of interest could be for example the survival of the patients since entry in a clinical study. Typically, some of the patients will not have experienced the event of interest at the time of the analysis; these patients will be considered as censored. So, we observe for $i = 1, \ldots, n$, $Z_i = \min(Y_i, C_i)$ with Y_i the time to event and C_i the censoring time, and the censoring indicator $\delta_i = 1_{\{Y_i \leq C_i\}}$. We also have for each patient a p-vector of covariates $\mathbf{X}_i = (X_{i1}, X_{i2}, \ldots, X_{ip})^t$. Assuming a random design, we therefore have an i.i.d. sample $(Z_i, \delta_i, \mathbf{X}_i)$, $i = 1, \ldots, n$, from $(Z, \delta, \mathbf{X})$. We assume that, conditional on $\mathbf{X}$, the censoring time and the time-to-event are independent (e.g., administrative censoring).

Modeling the relation between the suvival function $S(y|\mathbf{x}) = P(Y > y|\mathbf{X} = \mathbf{x})$ and the covariates is a key issue in survival analysis. Various models have been proposed, usually via modeling of the hazard function $\lambda(t|\mathbf{x})$. A

classical model is the Cox proportional hazards (PH) model (Cox, 1972)

$$\lambda(y|\mathbf{x}) = \lambda_0(y)\exp(\sum_{j=1}^{p}\beta_j x_j) \tag{1}$$

where $\lambda_0(y)$ is the baseline hazard function and $(\beta_1, \ldots, \beta_p)^t$ is the regression coefficients vector. In model (1) the regression coefficients are assumed to be constant over time. However, this is not always the case (e.g., the impact of a prognostic index value may decline over time, see for example Coradini et al. (2000)). Therefore extensions of model (1) having time-dependent regression coefficients have been proposed in the recent literature. An overview of such extensions is proposed in Buchholz and Sauerbrei (2011). The additive risk model (Klein and Moeschberger, 2003) and the proportional odds model (Shen, 1998) are also used as possible alternatives for the Cox model. Also for these models it is important to have extended versions allowing the regression coefficients to be time-dependent.

To study survival models with time-varying regression coefficients it is useful to develop, for all these models, a unified approach. Based on ideas by Nelder and McCullagh (1989) on generalized linear models, this can be done using an appropriate monotone link function $\phi : [0,1] \to \mathbb{R}$. The impact of the covariates on the conditional survival function can then be modeled as

$$\begin{aligned} \phi(S(y|\mathbf{x})) &= \beta_0(y) + \sum_{j=1}^{p}\beta_j(y)x_j \\ &= \beta_0(y) + \boldsymbol{\beta}^t(y)\mathbf{x} \end{aligned} \tag{2}$$

The choices for ϕ leading to the models mentioned above will be given in Section 2. Using maximum likelihood techniques, estimation of $(\beta_0(y), \boldsymbol{\beta}^t(y))$

has been studied by Jung (1996) and Subramanian (2001). More recently, Teodorescu et al. (2010) and Teodorescu and Van Keilegom (2010) use least squares estimation. Jung (1996) considers survival times that are right censored and covariates that are discrete and independent of the censoring times. Subramanian (2001) relaxes the independence assumption and allows a one-dimensional continuous covariate. Teodorescu et al. (2010) and Teodorescu and Van Keilegom (2010) study the case of right censored and/or left truncated time-to-event data and, for the covariates, allow a combination of discrete covariates and a one-dimensional continuous covariate. They also developed the asymptotic properties of the proposed estimators.

Model (2) generates a rich variety of conditional models. While previous papers focussed on the estimation techniques, one objective of this paper is to put more attention on the various conditional models covered by expression (2). Through (2) we can define a conditional survival function $S(y|\mathbf{x})$ and since $S(y|\mathbf{x}) = \phi^{-1}(\beta_0(y) + \boldsymbol{\beta}^t(y)\mathbf{x})$, the $\beta_j(.)$, $j = 0, \ldots, p$, should be chosen is such a way that the monotonicity and the boundary conditions required for a survival function are satisfied. For different choices of the link functions, we will give, in Section 2, the properties that $\beta_j(.)$ should have to guarantee that $\phi^{-1}(\beta_0(y) + \boldsymbol{\beta}^t(y)\mathbf{x})$ is a survival function. This investigation will reveal concrete families of conditional survival functions, resulting in an easy way to generate data from models having the structure of model (2), i.e., from models with time-dependent regression coefficients. Our second objective is to investigate the practical implementation and applicability of the proposed approach. To do so results on a bladder cancer database are used;

and we run simulations in various scenarios typically encountered in oncology.

In Section 3 we summarize the estimation method proposed in Teodorescu et al. (2010). They generate pseudo-values for $S(y|\mathbf{x})$ using a nonparametric kernel-based estimator (Beran, 1981), therefore for each separate event time (i.e., for each uncensored observation) a bandwidth needs to be selected. In Section 4 we propose a bandwidth selection algorithm that actually selects one bandwidth per group of event times, a procedure that results in computational efficiency. Using real data from a bladder cancer study we show how the proposed procedure works in practice.

Based on our findings for data generation and bandwidth selection, we perform in Section 5 a simulation study in the context of oncology clinical trials. These simulations challenge the method proposed in Teodorescu et al. (2010) for a model with regression coefficients that are constant in time and a model with a time-dependent regression coefficient for a continuous covariate. A general discussion and some further extensions of this work are briefly presented in Section 6.

2. Survival models with time-varying regression coefficients

Specific choices of the link function in model (2) give interesting classes of survival models with time-varying regression coefficients. We discuss three important classical examples. Our objective is not only to show the wide variety of possible model extensions but also to demonstrate that these results are useful for generating simulated survival data with time-dependent

coefficients.

Example 1: Cox proportional hazards model - With $\phi(s) = \ln(-\ln(s))$, model (2) can be written as

$$\ln(\Lambda(y|\mathbf{x})) = \beta_0(y) + \sum_{j=1}^{p} \beta_j(y) x_j$$

where $\Lambda(y|\mathbf{x})$ is the conditional cumulative hazard function and $\beta_0(y) = \ln(\Lambda_0(y))$ with $\Lambda_0(y)$ the baseline cumulative hazard function, or, equivalently,

$$\Lambda(y|\mathbf{x}) = \Lambda_0(y) \exp(\boldsymbol{\beta}^t(y)\mathbf{x}) \tag{3}$$

Model (3) clearly extends the classical Cox model $\Lambda(y|\mathbf{x}) = \Lambda_0(y) \exp(\boldsymbol{\beta}^t\mathbf{x})$ to time-dependent coefficients. Note that the way in which (3) extends the Cox model is different from the extension of Abrahamowicz et al. (1996) who consider an extension of form $\lambda(y|\mathbf{x}) = \lambda_0(y) \exp(\boldsymbol{\beta}^t(y)\mathbf{x})$. These two possible ways to extend the Cox model to a conditional survival model, i.e., introducing the time-dependency of the regression coefficients at the level of the cumulative hazard function or at the level of the hazard function, lead to different families of conditional survival functions. This will become clear from the specific examples given below.

Example 2: Proportional odds model - The logit link $\phi(s) = \ln(s/(1-s))$ leads to the logistic model

$$\ln\left(\frac{S(y|\mathbf{x})}{1 - S(y|\mathbf{x})}\right) = \beta_0(y) + \sum_{j=1}^{p} \beta_j(y) x_j \tag{4}$$

which, compared to the survival analysis proportional odds model studied for example by Murphy et al. (2007), allows to study time-dependent covariate effects.

Example 3: Additive hazards model - With $\phi(s) = -\ln(s)$, model (2) extends the additive risk model discussed in Martinussen and Scheike (2009) and in Chapter 10 of Klein and Moeschberger (2003). For this choice of the link function, we have

$$\Lambda(y|\mathbf{x}) = \beta_0(y) + \sum_{j=1}^{p} \beta_j(y) x_j \tag{5}$$

The classical additive risk model $\lambda(y|\mathbf{x}) = \alpha_0(y) + \sum_{j=1}^{p} \alpha_j x_j$ is a special case of (5) with $\beta_0(y) = \int_0^y \alpha_0(v)dv$ and $\beta_j(y) = \alpha_j y, j = 1, \ldots, p$.

Inference for models (3)-(5) is important; indeed fitting these models and comparing them with the fit obtained from the corresponding model having time-independent coefficients can provide evidence for the validity of the latter models (goodness-of-fit argument). Moreover, if the regression coefficients are indeed time-dependent, we need to incorporate this in our statistical analysis to arrive at sound inferential conclusions.

Note that models (3)-(5) can be rewritten as $S(y|\mathbf{x}) = \phi^{-1}(\beta_0(y) + \boldsymbol{\beta}^t(y)\mathbf{x})$. To guarantee that $S(y|\mathbf{x})$ is a well defined conditional survival function, i.e., $S(y|x)$ is, for a given x, decreasing in y with $S(0|x) = 1$ and $S(+\infty|x) = 0$, the choice of the link functions will put conditions on $\beta_j(.)$, $j = 0, \ldots, p$. For the case of a single covariate we collect in Table 1 sufficient conditions on

$\beta_0(.)$ and $\beta_1(.)$ for the link functions in Examples 1-3.

TABLE 1 about here.

Remark 1. For the $\ln(-\ln(.))$ link function (related to the Cox model) β_0 constant is not a possible choice since $\beta_0(y) = \ln \Lambda_0(y)$ with $\Lambda_0(.)$ the cumulative baseline hazard.

In the following discussion we show, for the link functions in Examples 1-3, how specific choices for $\beta_0(.)$ and $\beta_1(.)$ lead to extensions of well-known families of conditional survival functions and to new families. An important result emerging from this discussion is that it provides an easy way to generate or to simulate data from survival models with time-dependent regression coefficients. This finding is of major interest given that very little results are available on the generation of survival data with time-dependent coefficients.

Example 1: For the $\ln(-\ln(.))$ link, if we consider $\beta_0(y) = \ln \Lambda_0(y) = \ln(\lambda_0 y^{\beta_0})$ and $\beta_1(y) = \beta_1 \ln y$ we have

$$S(y|x) = \exp(-\lambda_0 y^{\beta_0+\beta_1 x}) \tag{6}$$

or

$$\lambda(y|x) = \lambda_0(\beta_0 + \beta_1 x) y^{(\beta_0+\beta_1 x-1)} \tag{7}$$

This is a conditional Weibull model with scale parameter λ_0 and shape parameter $\beta_0 + \beta_1 x$. Data can then be easily generated from model (7). Note

that this model is not contained in the class of models proposed by Abrahamowicz et al. (1996) and used also by Dunkler et al. (2010) to generate data with time dependent coefficients.

Example 2: For the *logit* link, we give three concrete families, each family is a log-logistic survival family.
For $\beta_0(y) = -\ln \beta_0$ and $\beta_1(y) = -\beta_1 \ln y$, we have

$$S(y|x) = (1 + \beta_0 y^{\beta_1 x})^{-1} \tag{8}$$

For $\beta_0(y) = -\beta_0 \ln y$ and $\beta_1(y) = -\ln \beta_1$, we have

$$S(y|x) = (1 + \beta_1^x y^{\beta_0})^{-1} \tag{9}$$

For $\beta_0(y) = -\beta_0 \ln y$ and $\beta_1(y) = -\ln \beta_1 - \beta_2 \ln y$, we have

$$S(y|x) = (1 + \beta_1^x y^{\beta_0 + \beta_2 x})^{-1} \tag{10}$$

Example 3: Take $\phi(s) = -\ln s$. For $\beta_0(y) = 0$ and $\beta_1(y) = -\beta_1 y$, we have

$$S(y|x) = \exp(-\beta_1 x y) \tag{11}$$

This is a conditional exponential model.
Another possibility would be $\beta_0(y) = 0$ and $\beta_1(y) = \beta_1 y^2$, for which we have

$$S(y|x) = \exp(-\beta_1 x y^2) \tag{12}$$

Remark 2. Since the above discussion serves the purpose of data generation only parmetric baseline hazards are considered.

3. Estimation of the time-dependent coefficients

Teodorescu et al. (2010) and Teodorescu and Van Keilegom (2010) obtain a least squares estimator $(\hat{\beta}_0(y), \hat{\beta}_1(y))^t$ for the time-dependent parameter $(\beta_0(y), \beta_1(y))^t$ and they show that, under regularity conditions, the asymptotic distribution is bivariate normal. They also propose a goodness-of-fit statistic for their time-varying coefficients model and show that the limit distribution of the proposed statistic is normal. For right censored data and for a single continuous covariate (they allow an extra discrete covariate) we extract from their work the inferential tools we need for the model we consider, i.e.,

$$\phi(S(y|x)) = \ln(-\ln(S(y|x))) = \beta_0(y) + \beta_1(y)x \tag{13}$$

The proposed estimation techniques, which can easily be adapted to more complex linear models, works as follows.

For a fixed value of y, we estimate $S(y|x_i)$, $i = 1, \ldots, n$, using the nonparametric Beran estimator (Beran, 1981; Van Keilegom et al., 2001):

$$\hat{S}_n(y|x) = \prod_{z_i \leq y, \delta_i = 1} \{1 - \frac{B_i(x,h)}{\sum_{j=1}^n I(z_j \geq z_i) B_j(x,h)}\} \tag{14}$$

where $B_i(x,h)$ are the Nadaraya-Watson weights

$$B_i(x,h) = \frac{K(\frac{x-x_i}{h})}{\sum_{j=1}^n K(\frac{x-x_j}{h})}$$

with $K(.)$ a known density function (kernel), and $h = h_n \to 0$ a bandwidth sequence. Following the work of Van Keilegom et al. (2001), we use the biquadratic kernel $K(z) = (15/16)(1-z^2)^2 I(|z| \leq 1)$. Note that if x is a

categorical variable rather than a continuous one, the Beran estimator can be replaced by the Kaplan-Meier estimator computed for each category.
We compute $\phi(\hat{S}_n(y|x_i))$, $i = 1, \ldots, n$, and

$$\hat{\boldsymbol{\phi}}(y) = (\phi(\hat{S}_n(y|x_1), \ldots, \phi(\hat{S}_n(y|x_n))^t \tag{15}$$

the vector of "pseudo-responses". We then fit model (13) using the components of $\hat{\boldsymbol{\phi}}(y)$ as responses. At this step, any linear model fitting strategies could be used. We apply the classical weighted least squares method to compute the estimators of $\beta_0(y)$ and $\beta_1(y)$, i.e.,

$$\begin{pmatrix} \hat{\beta}_0(y) \\ \hat{\beta}_1(y) \end{pmatrix} = (\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}\hat{\boldsymbol{\phi}}(y) \tag{16}$$

and $\mathbf{W} = \text{diag}(w(x_1), \ldots, w(x_n))$, with $w(.) = 1_{\{x \in I\}}$ some weight function. This weight function downweights x_i values close to the boundaries of the range of the covariate values. Indeed, the Beran estimator is known to provide less accurate estimates at boundary points of the set of covariate values leading therefore to heteroscedasticity in our linear modeling.

By repeating this for all possible y we obtain an estimate $(\hat{\beta}_0(y), \hat{\beta}_1(y))^t$ whose value can vary over time. The functions $\hat{\beta}_0(y)$ and $\hat{\beta}_1(y)$ are step functions, because $\hat{S}_n(y|x)$ is a step function, and therefore it is sufficient to consider only y-values that are event times.
Using $(\hat{\beta}_0(y), \hat{\beta}_1(y))^t$ we can also obtain the following model-based estimator for the survival function

$$\tilde{S}(y|x) = \phi^{-1}(\hat{\beta}_0(y) + \hat{\beta}_1(y)x)$$

The fact that the monotonicity of $\tilde{S}(y|x)$ is not guaranteed is discussed and resolved using isotonic regression in Section 4.3.

4. Practical implementation

4.1. Case study: Bladder Cancer Data

We consider a pooled database of seven EORTC trials in bladder cancer patients. These trials were all run in the nineties to study the impact of intravesical chemotherapy after transurethral resection on the risk of recurrence and disease progression for patients with superficial bladder cancer. A total of 2596 stage Ta/T1 bladder cancer patients are included in this database. More details can be found in Sylvester et al. (2006), Legrand et al. (2009), and López-de Ullibarri et al. (2012).

Following López-de Ullibarri et al. (2012) we consider time to first recurrence, also called disease-free interval (DFI), as outcome variable of interest (censoring patients without recurrence at the date of last follow-up cystoscopy or death) and we restrict our analyses to the subset of 1405 patients having received an effective treatment (according to the results of these seven trials). In this group of patients, a total of 618 events (recurrences) were observed and the median DFI is 3.5 years. In this paper, we study further the continuous prognostic score, SC, defined by López-de Ullibarri et al. (2012) based on the following risk factors: number of tumors, tumor size in centimeters, prior recurrence rate (per year), T category, presence of concomitant carcinoma in situ and grade. This prognostic score is close to the one originally developed by Sylvester at al. (2006) on the same data, except that conti-

nous risk factors used to compute this prognostic score are not discretized. The theoretical values for SC range between 0 and 17, however in our data SC takes value between 0.16 and 12.87 with lower values corresponding to better prognosis (López-de Ullibarri et al., 2012, Figure 2). All the variables included in this prognsotic score, and therefore the prognostic score itself, are measured at baseline and a question of interest is whether the effect of this score is constant over time or not.

4.2. Bandwidth selection

To estimate $(\beta_0(y), \beta_1(y))^t$ we need, for each event time y, the Beran estimator of $S(y|x_i)$. Hence, a bandwidth h needs to be selected. Based on the idea of Teodorescu et al. (2010), we implemented a bootstrap procedure to select, for a given event time y, an appropriate bandwidth h taken from a grid $\{h_1, h_2, \ldots, h_r\}$ of possible bandwidths. These $h_j, j = 1, \ldots, r$, are arbitraily chosen in $[0, L/2]$ where L is the upper bound of the covariate values.

For y fixed and h_j fixed, the bootstrap algorithm given in the Appendix provides us with a value for

$$MSE^{\star}_{h_j}(y) = \frac{1}{B}\sum_{l=1}^{B}\left[\left(\hat{\beta}^{\star}_{0,lh_j}(y) - \hat{\beta}_{0,g_j}(y)\right)^2 + \left(\hat{\beta}^{\star}_{1,lh_j}(y) - \hat{\beta}_{1,g_j}(y)\right)^2\right]$$

where B is the number of bootstrap samples, $\hat{\beta}^{\star}_{i,lh_j}(y), i = 0, 1$, are the bootstrap estimators of $\beta_i(y)$ obtained with bandwidth h_j and $\hat{\beta}_{0,g_j}(y)$ are the estimators of $\beta_i(y)$ obtained on our data but using a pilot bandwidth g_j. As discussed by Iglesias-Pérez and Gonzales-Manteiga (2003), the use of such a pilot bandwidth g_j aims to ensure that the bootstrap bias and the bootstrap

variance are asymptotically good estimators for the bias and the variance.

We can then select the bandwidth that minimizes this mean squared error (MSE), i.e.,

$$h^{\star}_{OPT}(y) = arg_{\{h_1,\ldots,h_r\}} min MSE^{\star}_{h_j}(y)$$

This idea can be repeated for each y.

Another possibility, as suggested by Teodorescu et al. (2010), is to consider a global bandwidth h that minimizes

$$MISE^{\star}_{h_j} = \int_a^b MSE^{\star}_{h_j}(y)dy \tag{17}$$

for a specific interval $[a, b]$ satisfying the conditions in Teodorescu et al. (2010).

However, these two approaches are very time consuming and, from our experience, it appears that *(i)* selecting a different bandwidth for each event time is not necessary as most of the selected bandwiths will be very close for several adjacent timepoints, and *(ii)* selecting a global bandwidth might be misleading as in most cases we observed a dramatic drop in the bandwidth selected for the earlier event times and the later event times. A more practical approach is to split the ordered event times in groups and to select one bandwidth per group. This pragmatic approach has been applied to our data (and in our simulations), considering $B = 500$ bootstrap samples. Possible values of the bandwidth considered were between 0.5 and 5 with span 0.5. We used 5 groups of about 120 events each. In each group, the MSE

was computed for each lower and upper event times as well as for 3 (equally spaced) inner timepoints. The optimal bandwidth for each group was then obtained by minimizing the average MSE for these 5 event times. Table 2 represents the average MSE obtained for each group and each potential value of the bandwidth as well as the final choice made to analyse the data. To test the stability of our procedure, we have repeated it 5 times (with $B = 200$) and we have found each time very close results. We could also have selected an optimal bandwidth obtained by minimizing the average MSE of $\hat{\beta}_1$ as we are mainly interested in this parameter. This however led to very similar results.

TABLE 2 about here.

4.3. Practical implementation: isotonic regression, weighted regression, and computational issues in $\ln(-\ln(.))$ *transform*

The estimation algorithm as well as the boostrap-based bandwidth selection algorithm discussed above lead to some typical issues to be solved, namely the non monotonicity of the estimated survival function, the choice of the weight function for the weighted regression, and the computation of the $\ln(-\ln(.))$ transform for values of 0 or 1.

First, the possible non monotonicy of the model-based conditional survival (and censoring) distribution was solved using the idea of isotonic regression (de Leeuw J. et al., 2009). Indeed, the naive solution proposed by Teodorescu

et al. (2010) which consists of keeping the estimated value constant up to the time it starts decreasing again (when going from the start to the end of the curve) is not satisfying as we observed frequently a rather steep drop in the curve at early event times. For the weighted regression, we followed the idea of Teodorescu et al. (2010) and considered as weight function $w(x) = 1_{\{x \in [a,b]\}}$ with a and b respectively the 2.5% and the 97.5% percentile of the ordered covariate values. Finally, the fact that the $\ln(-\ln(.))$ transform cannot be computed when $S(y|x)$ equals 0 or 1, was solved by adding/removing the smallest quantity required for R to be able to compute this transform.

For the (technical) details on these issues we refer to our R code (available from the first author).

4.4. Results

Ignoring the potential time-dependency in the effect of SC on DFI and fitting a classical Cox proportional hazards model to our data, leads to an estimated coefficient of 0.1651 corresponding to a HR of 1.180 (95%CI:[1.150, 1.210]). In Table 3, the values for $\hat{\beta}_0(y)$ and $\hat{\beta}_1(y)$ and their standard errors obtained when fitting model (2) are given for a subset of 5 timepoints (mid-point of the periods defined to choose the optimal bandwidth). The standard errors obtained from the weighted linear regression are obtained considering the pseudo-responses as independent and therefore underestimate the real standard errors. We therefore use the empirical standard errors computed from the bootstrap replications performed for bandwidth selection.

TABLE 3 about here.

Figure 1 represents the estimated values of $\beta_0(y)$ and $\beta_1(y)$ obtained when fitting model (2) for all event times; in the right panels, information is rectricted to time 0-1000 days. As expected, $\hat{\beta}_0(y)$, estimating $\ln(\Lambda_0(y))$, is increasing with time. More interestingly, $\hat{\beta}_1(y)$ also seems to vary with time, with an inverted "U-shape". For the first 250 days, the coefficient $\hat{\beta}_1(.)$ is increasing, then reaching some kind of plateau and then later starting to decrease.

FIGURE 1 about here.

In Figure 2, $\exp(\hat{\beta}_1(y))$ represents the estimated ratio of cumulative hazard functions for patients with a SC value of $v+1$ compared to a value of v. Clearly, patients with higher value of SC at baseline remain with an increased cumulative hazard function over time, but with an increasing ratio over the first half year, reaching then a plateau with an about 20-25% higher cumulative hazard function up to a bit less than two years and decreasing then to an about 10% higher risk. This indicates an increased risk with increasing value of the prognostic index, however this increased risk tends to get smaller over time. This can be seen also on the plot of the survival curves (Figure 3) obtained from $\phi^{-1}(\hat{\beta}_0(y) + \hat{\beta}_1(y)x)$ for $x = 1$, 5, 10 and 15; as explained earlier, isotonic regression is used to impose monotonocity of these curves.

FIGURE 2 about here.

FIGURE 3 about here.

5. Simulations

5.1. Data Generation

For data generation, we use the results obtained in Section 2. Considering the $\ln(-\ln(.))$ link function, a Weibull baseline hazard, i.e., $\beta_0(y) = \ln(\Lambda_0(y)) = \ln(\lambda_0 y^{\beta_0})$ and $\beta_1(y) = \ln(y^{\beta_1})$ yield, as shown previously, a conditional Weibull distribution for the event times with scale parameter λ_0 and shape parameter $\beta_0 + \beta_1 x$. Hence, data are generated as follows:

1. Generate covariate values x_i, $i = 1, \ldots, n$, from $X \sim f_X(.)$, e.g., from $X \sim U[a, b]$ uniform on [a,b].
2. Choose appropriate values for λ_0, β_0, and β_1. These choices can be based on medical evidence, solving $S(y|x) = q$ for different medically relevant values of y, x, and q (see Table 4).
3. Generate event times y_i, $i = 1, \ldots, n$ from $Y_i \sim W(\lambda_0, \beta_0 + \beta_1 x_i)$, a Weibull distribution with scale parameter λ_0 and shape parameter $\beta_0 + \beta_1 x_i$.
4. The assumed censoring mechanism is staggered entry, i.e., with A the total duration of accrual, and F the duration of follow-up time after last patient entry, c_i, $i = 1, \ldots, n$ is obtained from $i \times A/n + F$.
5. Compile (z_i, δ_i, x_i), $i = 1, \ldots, n$, $z_i = \min(y_i, c_i)$ and $\delta_i = 1_{\{y_i \leq c_i\}}$.

In this paper, simulations are run in the setting of the bladder cancer clinical trials presented in Section 4. This procedure could clearly be generalized to other settings. Our continuous covariate mimicking SC in the previous

section, is assumed to take values uniformly in $[0, 20]$. We will consider six different scenarios. In the two first ones, a constant effect is assumed for the effect of x and event times are then simulated from a Weibull distribution with scale parameter $\lambda_0 \exp(\beta_1 x)$ and shape parameter β_1. For the four latter ones, effect of x is assumed to be varying with time. Various scenarios are summarized in Table 4 considering different prognoses and different effects of x over time. A nice feature of the proposed approach to simulate data is that choice of the parameters can be made to mimick some medically relevant situation. We indeed obtained our simulations parameters by solving a system of equations $S(y|x) = q$, for given choice of y, x and q and with $S(.|.)$ given by (6). While three equations are sufficient to specify the parameters, we used four equations to better control the decay at the start of the survival curve and we obtain an approximated solution of the overdetermined set of equations using the function BBsolve in R (Varadhan and Gilbert, 2009). For example scenario 3 is obtained considering a patient population with survival percentages at 10 years of 60%, 75%, and 30% for patients with prognostic index values of respectively 10, 5 and 15; while patients with a prognostic index value of 10 are expected to have a survival at 1 year of 95%.

TABLE 4 about here.

To get a better idea of the data simulated, Figure 4 plots typical simulated KM survival curves for $x = 5, 10$ and 15 for each scenario (considering no censoring). For each scenario, we generate 500 datasets of size $n = 1000$. Considering an accrual period of 5 years and a follow-up period of 6 years

leads to administrative censoring rates of about 56%, 36%, 57%, 24%, 39% and 16% for scenarios 1 to 6. The bandwidth choice algorithm described in Section 4.2 was run on the first generated dataset for each scenario and optimal choice was then applied to all datasets for this setting. All the simulations were performed in R, the code is available from the first author.

FIGURE 4 about here.

5.2. Results

The estimated values of $\beta_0(y)$ and $\beta_1(y)$ for all timepoints and for each simulated dataset are represented on Figures 5 and 6 respectively. On these plots, the dark lines represent the true curves as computed from simulations parameters. For sake of clarity, only one plot for the case of time-inependent coefficient (scenario 2) and one for the case of time-depedent coefficient (scenario 5) are presented. As expected, we observe, in all scenarios, a large variability in our estimation at the earliest timepoints. At these timepoints, estimated values of $\hat{S}(y|x)$ are mostly close to 0 resulting in very variable peudo-responses due to the $\ln(-\ln(.))$ transform of these values. We also observe some over/under estimation of $\beta_0(y)$ and $\beta_1(y)$ respectively at later timepoints, when the number of patients at risk become low and only little information is available for estimation. Nevertheless, our estimation method clearly picks up the time-dependency of the coefficients.

FIGURE 5 about here.
FIGURE 6 about here.

This is even clearer on Figure 7 and Figure 8. There, for the six scenarios, and for six different timepoints (500, 1000, 1500, 2000, 2500, and 3000), we use boxplots to summarize, for the 500 generated datasets, the estimates obtained for $\beta_0(y)$ and $\beta_1(y)$. In these boxplots, the cross represents the true value of the parameter. The estimation is unbiased for most timepoints in each scenario, whether simulating the data with a fixed covariate effect (top panels) or a time-varying effect (middle and bottom panels). As one can expect, the method is however not reliable at the later timepoints when the number of patients still at risk becomes too low. This is particularly visible in scenario 6, where patients have a very poor prognosis and very few patients remains at risk after time 1000.

FIGURE 7 about here.
FIGURE 8 about here.

6. Discussion

The need to extend survival models to time-dependent coefficients has now been ackowledged by several authors. When survival models are used in medical applications to model the impact of prognostic factors or treatment variables on a time-to-event outcome, several examples have indeed shown time-varying effects of such factors. For example, in oncology, Coradini et al. (2000) have shown that the effect of oestrogen receptor and tumor size is varying over time. Often, extension to time-dependent coefficients is done starting from the Cox proportional hazards model and introducing therefore time-varying effects at the level of the hazard function, considering a prede-

fined function of time, piecewise constant effects, fractional polynomials or splines (Buchholz and Sauerbrei, 2011). In this paper, we consider an extension to time-dependent coefficients via a transformation model. Following the idea of generalized linear models, the effect of the covariates is modeled on the suvival function via a link function. This approach has the advantage of covering a broad range of models via a very general formulation of these models.

We build further on the work of Teodorescu et al. (2010); Teodorescu and Van Keilegom (2010) who proposed an estimation method based on least squares for this class of models. First, we explore the various models included in this class, and show that by considering an appropriate choice of the link function we obtain extensions of the various well-known models (Cox PH model, proportional odds model and additive risk model). Interestingly, several of these extensions are different from those previously proposed. For example, we focus in our real data analysis and in our simulations on one of these models, which is actually an extension of the Cox model but considering now the time-dependency of the coefficients at the level of the cumulative baseline hazard and not at the level of the hazard function. The proposed models therefore offer a different modeling strategy. Studying various models included in this class, our findings provide an easy way to generate data with time-dependent coefficients.

We also address several important issues encountered in the practical implementation of this approach. Among these, the bandwidth selection is a

crucial issue, as a nonparametric estimator is used for the survival of each patient at each timepoint. The original approach proposed by Teodorescu et al. (2010); Teodorescu and Van Keilegom (2010), considering a bootstrap procedure applied at each event time, quickly appeared unrealistic in terms of computation time. We propose a computational effective alternative that has similar accuracy.

Our simulations, performed in the context of bladder cancer, show that this approach can be used as a diagnostic tool to detect time-varying effect of a continuous covariate. Prognostic index, which are typically continuous variables, are based on variables measured at baseline with the objective to provide information on the outcome of the patients. When these patients are in follow-up, it is not natural to assume that the effect of the prognostic index is constant over time (considering eventual change in therapy, evolution of the disease, discharge of the patient, ...). It is therefore important to check for presence/absence of time dependency. Indeed, ignoring this time dependency may lead to wrong model specifications and therefore wrong conclusions. We illustrate our ideas considering a continuous prognostic score developed in bladder cancer, and show that indeed the effect of this score is to decreasing over time.

Other authors have been working on transformation models (Massonnet et al., 2008; López-de Ullibarri et al., 2012). However the ideas behind the estimation techniques are much more complex compared to our approach. Also, as will be shown in forthcoming work, the proposed method is also

useful to study models with time-varying frailties.

Acknowledgment

Research supported by IAP research network grant nr. P7/06 of the Belgian government (Belgian Science Policy). C. Legrand is supported by the the contract 'Projet d'Actions de Recherche Concertées' (ARC) 11/16-039 of the 'Communauté francaise de Belgique', granted by the 'Académie universitaire Louvain'. M. Munda is supported by a F.R.I.A. fellowship.
The authors thank the European Organization for Research and Treatment of Cancer for permission to use the data from EORTC studies 30781, 30782, 30791, 30831, 30832, 30845, 30863 for this research. The contents of this publication and methods used are solely the responsibility of the authors and do not necessarily represent the official views of the EORTC Headquarters.

References

References

Abrahamowicz, M., Mac Kenzie, T., Esdaile, J., 1996. Modeling and hypothesis testing with application in lupus nephritis. Journal of the American Statistical Association 91, 1432–1439.

Beran, R., 1981. Nonparametric regression with randomly censored survival data. Technical Report, University of California, Berkeley.

Buchholz, A., Sauerbrei, W., 2011. Comparison of procedures to assess nonlinear and time-varying effects in multivariable models for survival data. Biometrical Journal 53, 308–331.

Coradini, D., Daidone, M., Boracchi, P., Biganzoli, E., Oriana, S., G., B., Pellizarro, C., Tomasic, G., Di Fronzo, G., Marubini, E., 2000. Time-dependent relevance of steroid receptors in breast cancer. Journal of Clinical Oncology 18, 2702–2709.

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

de Leeuw J., Hornik, K., Mair, P., 2009. Isotone optimization in r: Pool-adjacent-violators algorithm (pava) and active set methods. Journal of Statistical Software 32, 1–24.

Dunkler, D., Schemper, M., Heinze, G., 2010. Gene selection in microarray survival studies under possibly non-proportional hazards. Bioinformatics 26, 784–790.

Iglesias-Pérez, C., Gonzales-Manteiga, W., 2003. Bootstrap for the conditional distribution function with truncated and censored data. The Annals of the Institute of Statistical Mathematics 55, 331–357.

Jung, S., 1996. Regression analysis for long-term survival rate. Biometrika 83, 227–232.

Klein, J., Moeschberger, M., 2003. Survival Analysis. Techniques for Censored and Truncated Data. Second Edition.

Legrand, C., Duchateau, L., Janssen, P., Ducrocq, V., Sylvester, R., 2009. Validation of prognostic indices using the frailty model. Lifetime Data Analysis 15, 59–78.

López-de Ullibarri, I., Janssen, P., Cao, R., 2012. Continuous covariate frailty models for censored and truncated clustered data. Journal of Statistical Planning and Inference, 142, 1864-1877.

Martinussen, T., Scheike, T., 2009. Covariate selection for the semiparametric additive risk model. Scandinavian Journal of Statistics 36, 602–619.

Massonnet, G., Janssen, P., Burzykowski, T., 2008. Fitting conditional survival models to meta-analytic data by using a transformation toward mixed-effects models. Biometrics 64, 834–842.

Murphy, S., Rossini, A., Van Der Vaart, A., 2007. Maximum likelihood estimation in the proportional odds model. Annals of the Institute of Statistical Mathematics 92, 545–574.

Nelder, J., McCullagh, P., 1989. Generalized Linear Models, Second Edition.

Shen, X., 1998. Proportional odds regression and sieve maximum likelihood estimation. Biometrika 85, 165–177.

Subramanian, S., 2001. Parameter estimation in regression for long-term survival rate from censored data. Journal of Statistical Planning and Inference 99, 211–222.

Sylvester, R., Van der Meijden, A., Oosterlinck, W., Witjes, J., Bouffioux, C., Denis, L., Newling, D., Kurth, K., 2006. Predicting recurrence and progression in individual patients with stage tat1 bladder cancer using eortc risk tables: a combined analysis of 2596 patients from seven eort trials. European Journal of Urology 49, 466–477.

Teodorescu, B., Van Keilegom, I., 2010. Goodness-of-fit test for generalized conditional linear models left truncation and right censoring. Journal of Nonparametric Statistics 22, 547–566.

Teodorescu, B., Van Keilegom, I., Cao, R., 2010. Generalized time-dependent conditional linear models under left truncation and right censoring. The Annals of the Institute of Statistical Mathematics 62, 465–485.

Van Keilegom, I., Akritas, M., Veraverbeke, N., 2001. Estimation of the conditional distribution in regression with censored data: a comparative study. Computational Statistics and Data Analysis 35, 487–500.

Varadhan, R., Gilbert, P., 2009. Bb: An r package for solving a large system of nonlinear equations and for optimizing a high-dimensional nonlinear objective function. J. Statistical Software 32.

$\phi(s)$	$S(y\|x)$	Constraints	
		$\beta_0(.)$	$\beta_1(.)$
$\ln(-\ln(s))$	$\exp(-\exp(\beta_0(y)+\beta_1(y)x))$	$[0,+\infty) \nearrow (-\infty,+\infty)$	constant
		$[0,+\infty) \nearrow (-\infty,+\infty)$	$[0,+\infty) \nearrow (-\infty,+\infty)$
$\ln\left(\frac{s}{1-s}\right)$	$\frac{\exp(\beta_0(y)+\beta_1(y)x)}{1+\exp(\beta_0(y)+\beta_1(y)x)}$	constant	$[0,+\infty) \searrow (-\infty,+\infty)$
		$[0,+\infty) \searrow (-\infty,+\infty)$	constant
		$[0,+\infty) \searrow (-\infty,+\infty)$	$[0,+\infty) \searrow (-\infty,+\infty)$
$-\ln(s)$	$\exp(-\beta_0(y)-\beta_1(y)x)$	constant	$[0,+\infty) \nearrow [0,+\infty)$
		$[0,+\infty) \nearrow [0,+\infty)$	constant
		$[0,+\infty) \nearrow [0,+\infty)$	$[0,+\infty) \nearrow [0,+\infty)$

Table 1: Sufficient constraints imposed on the shape of the function $\beta_0(.)$ and $\beta_1(.)$ for the link functions in Examples 1-3.

Period	I	II	III	IV	V
Time interval	21-98	98-184	184-350	350-669	669-3311
MSE (h=0.5)	210.4814	0.0591	0.0239	0.0151	1.2847
MSE (h=1)	16.6066	0.0425	0.0194	0.0137	0.3456
MSE (h=1.5)	24.3846	0.0299	0.0170	**0.0114**	0.2911
MSE (h=2)	12.9192	0.0283	**0.0145**	0.0118	0.2282
MSE (h=2.5)	**9.8706**	**0.0282**	0.0157	0.0123	0.1526
MSE (h=3)	12.6385	0.0335	0.0198	0.0146	0.1150
MSE (h=3.5)	24.4075	0.0287	0.0184	0.0149	0.0809
MSE (h=4)	49.9189	0.0308	0.0185	0.0147	0.0645
MSE (h=4.5)	77.5315	0.0378	0.0269	0.0225	**0.0360**
MSE (h=5)	132.8887	0.0346	0.0294	0.0232	0.0561
Optimal h	h=2.5	h=2.5	h=2	h=1.5	h=4.5

Table 2: Bladder cancer database: Choice of the optimal bandwidth

$\hat{\beta}_0$	SE($\hat{\beta}_0$)	$\hat{\beta}_1$	SE($\hat{\beta}_1$)	Mean($\hat{\beta}^\star_{0b}$)	SE($\hat{\beta}^\star_{0b}$)	Mean($\hat{\beta}^\star_{1b}$))	SE($\hat{\beta}^\star_{1b}$)
Period I (h=2.5) - Time = 60							
-5.1745	0.0136	0.0916	0.0028	-5.6381	1.3978	0.1691	0.3040
Period II (h=2.5) - Time = 141							
-2.9080	0.0061	0.2061	0.0012	-2.8106	0.1624	0.1828	0.0263
Period III (h=2) - Time = 267							
-2.1748	0.0052	0.2011	0.0011	-2.1030	0.1106	0.1841	0.0190
Period IV (h=1.5) - Time = 510							
-1.7685	0.0065	0.2044	0.0013	-1.7455	0.1038	0.1967	0.0187
Period V (h=4.5) - Time = 1990							
-0.7146	0.0015	0.1193	0.0003	-0.5086	0.0740	0.0687	0.0133

Table 3: Bladder cancer database: Estimated values of β_0 and β_1 at various timepoints; SE($\hat{\beta}_0$) and SE($\hat{\beta}_1$) are obtained from linear regression at each timepoint but have to be corrected for dependency between the pseudo-observations. The last four columns are obtained from the 500 bootstrap replications obtained for the choice of the optimal bandwidth.

	Simulations parameters	Obtained by solving
Scenario 1	$\lambda_0 = 0.0008$	$S(10\|10) = 0.60$
Constant coefficient	$\beta_0 = 0.5862$	$S(10\|5) = 0.75$
Good overall prognosis	$\beta_1 = 0.1714$	$S(10\|15) = 0.30$
		$S(1\|10) = 0.90$
Scenario 2	$\lambda_0 = 0.0142$	$S(2\|10) = 0.70$
Constant coefficient	$\beta_0 = 0.4890$	$S(2\|5) = 0.90$
Poor overall prognosis	$\beta_1 = 0.0366$	$S(2\|15) = 0.60$
		$S(5\|10) = 0.10$
Scenario 3	$\lambda_0 = 0.0016$	$S(10\|10) = 0.60$
Time-varying coefficient	$\beta_0 = 0.5252$	$S(10\|5) = 0.75$
Good overall prognosis	$\beta_1 = 0.0186$	$S(10\|15) = 0.30$
		$S(1\|10) = 0.95$
Scenario 4	$\lambda_0 = 0.0483$	$S(2\|10) = 0.50$
Time-varying coefficient	$\beta_0 = 0.3024$	$S(2\|5) = 0.60$
Poor overall prognosis	$\beta_1 = 0.0133$	$S(2\|15) = 0.30$
		$S(5\|10) = 0.10$
Scenario 5	$\lambda_0 = 0.0016$	$S(10\|10) = 0.60$
Time-varying coefficient	$\beta_0 = 0.5252$	$S(10\|5) = 0.75$
Good overall prognosis	$\beta_1 = 0.0279$	$S(10\|15) = 0.30$
		$S(1\|10) = 0.95$
Scenario 6	$\lambda_0 = 0.0483$	$S(2\|10) = 0.50$
Time-varying coefficient	$\beta_0 = 0.3024$	$S(2\|5) = 0.60$
Poor overall prognosis	$\beta_1 = 0.0200$	$S(2\|15) = 0.30$
		$S(5\|10) = 0.10$

Table 4: Choice of parameters value for the simulations - different scenarios. For scenarios 1 and 2, $\beta_1(y) = \beta_1$ (constant coefficient) while for scenarios 3 to 6, $\beta_1(y) = \beta_1 \ln(y)$

.

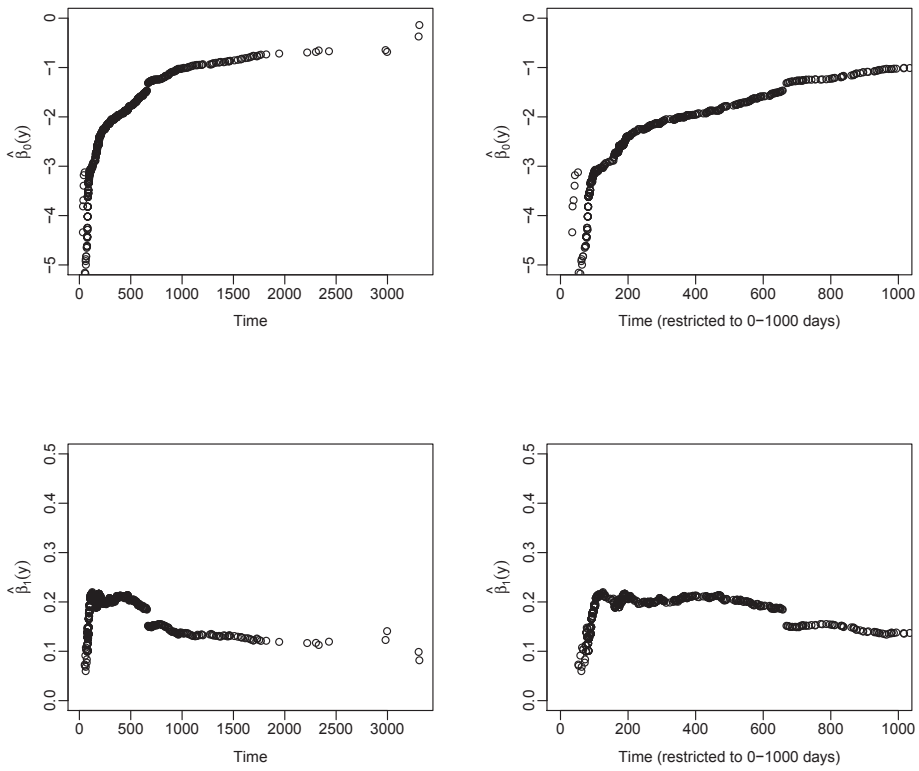

Figure 1: Estimated values for $\beta_0(y)$ (up) and $\beta_1(y)$ (down); right panels are restricted to timepoints 1-1000

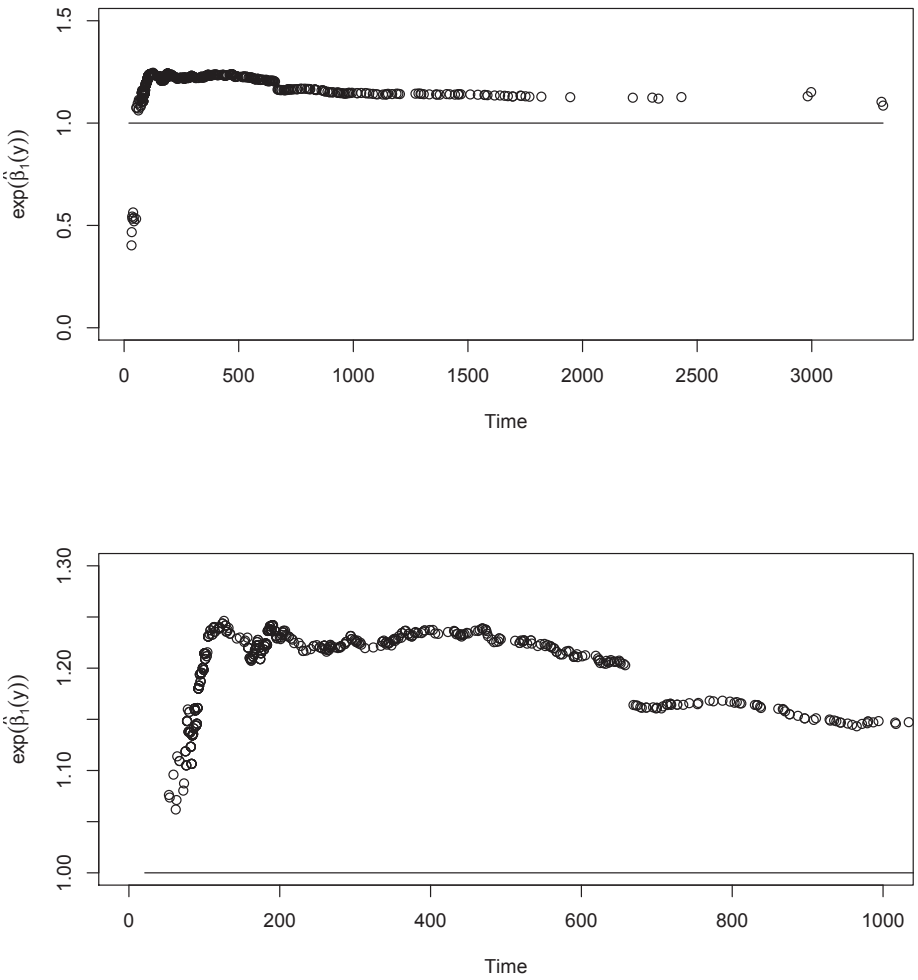

Figure 2: Estimated ratio of cumulative hazard functions for patients with a SC value of $v + 1$ compared to a value of v; bottom panel is restricted to timepoints 1-1000

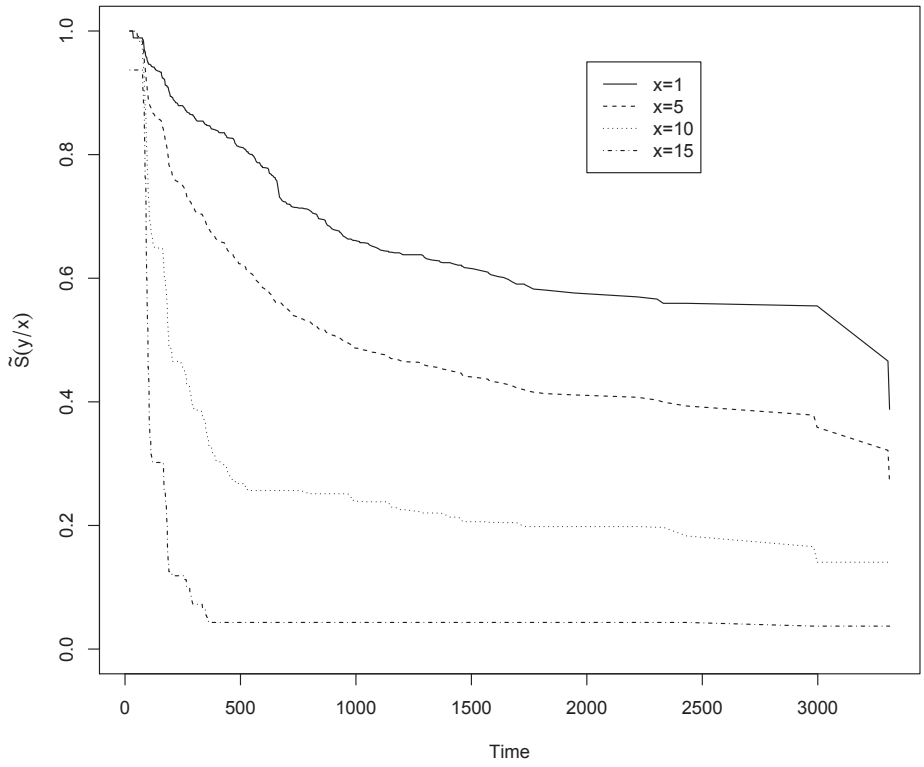

Figure 3: Estimated survival curves values from $\phi^{-1}(\hat{\beta}_0(y) + \hat{\beta}_1(y)x)$ for $x = 1$, 5, 10 and 15

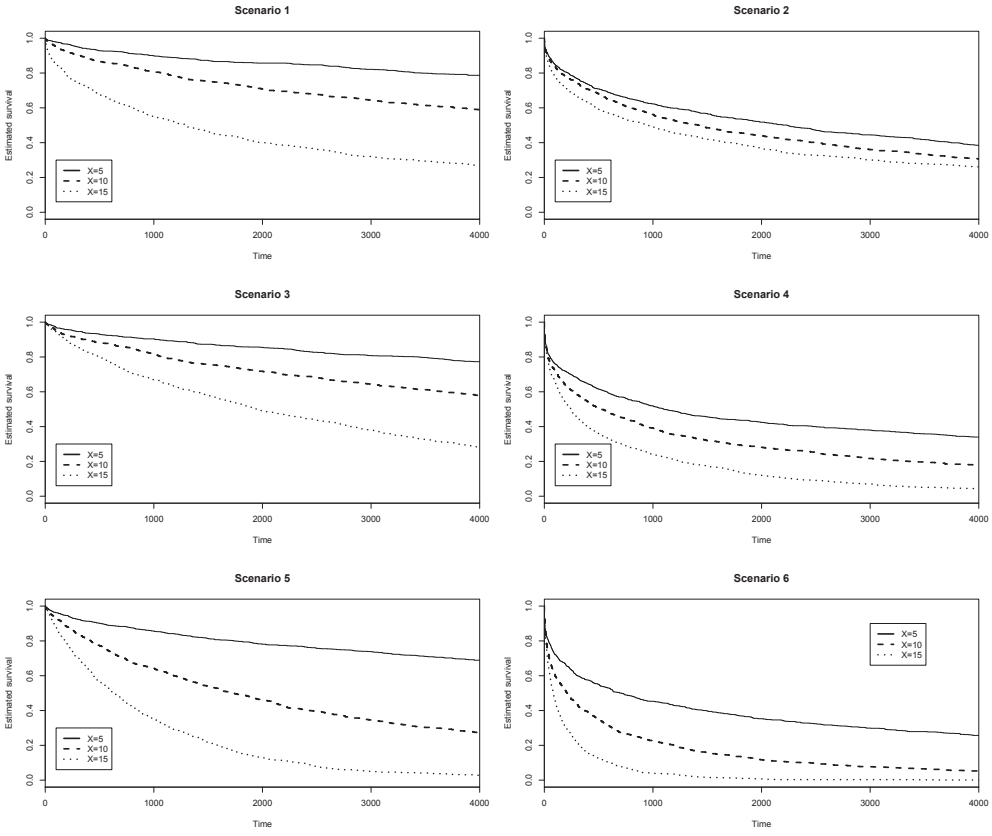

Figure 4: Typical simulated survival curves for $x = 5$ (—), 10 (- - -) and 15 ($\cdots$) and for each scenario considered in the simulations. These curves have been simulated assuming no censoring.

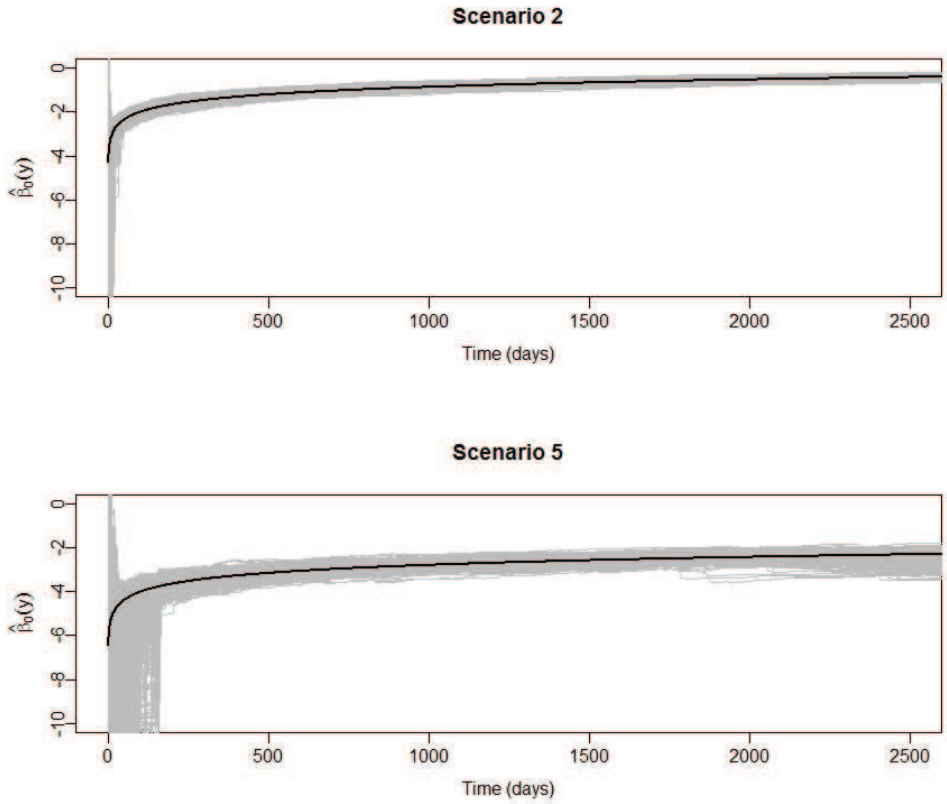

Figure 5: Estimated values for $\beta_0(y)$ at all event times for all simulated datasets for scenarios 2 (top) and 5 (bottom). In each graph, the dark line represents the true values used for simulations. For clarity of representation the Y-axis has been truncated to [-10,0] and time axis to [0,2500].

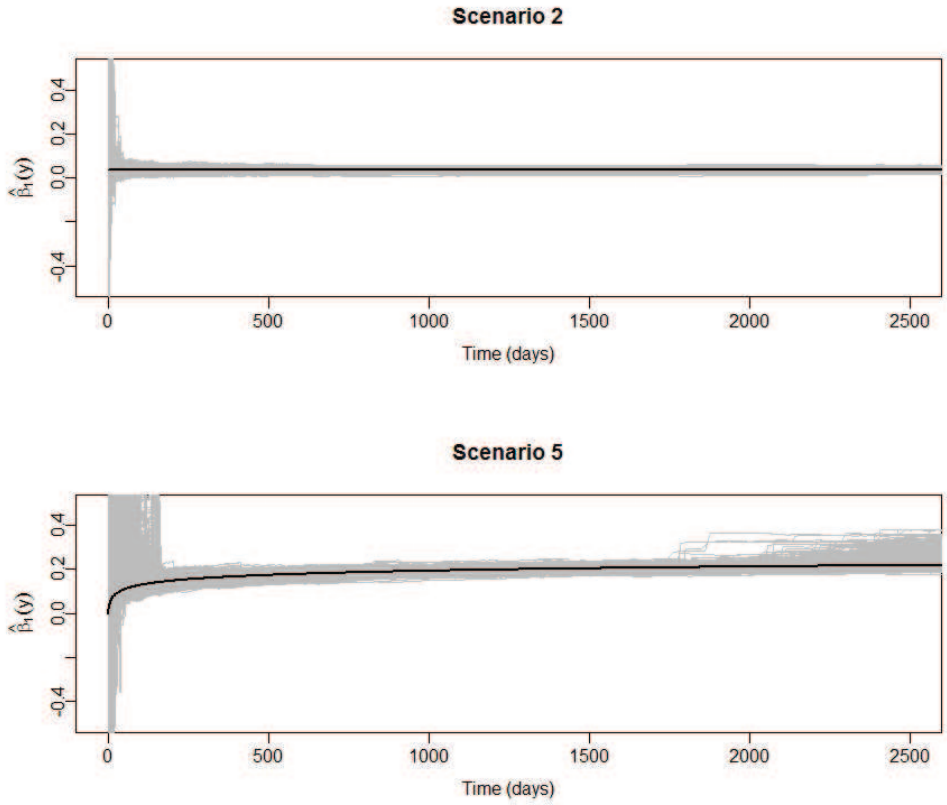

Figure 6: Estimated values for $\beta_1(y)$ at all event times for all simulated datasets for scenarios 2 (top) and 5 (bottom). In each graph, the dark line represents the true values used for simulations (effect assumed constant for scenario 2 and time-varying for scenario 5). To illustrate more clearly the time-dependency in scenario 5, Y-axis has been truncated to [-10,0] and time axis to [0,2500].

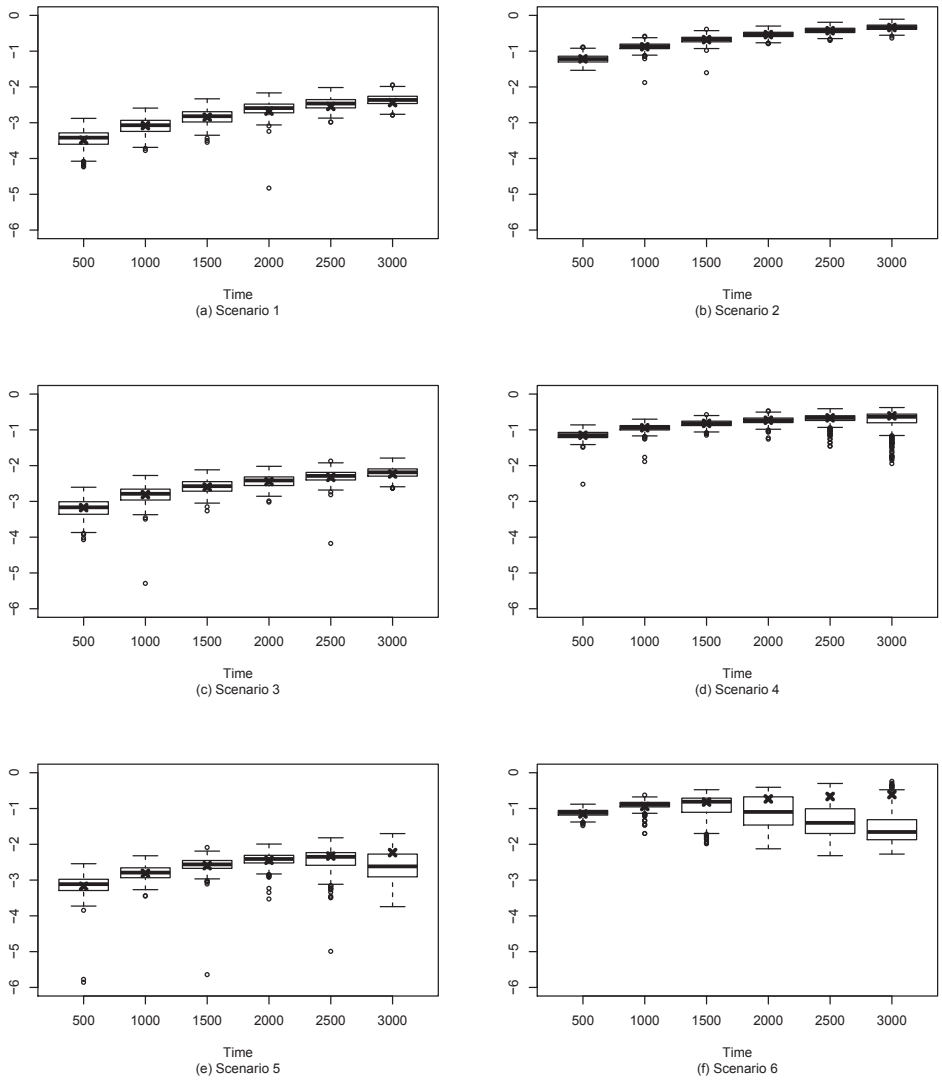

Figure 7: Results of the simulation for β_0: boxplots of the 500 estimated values obtained in each scenario, a cross represents the true value of β_0.

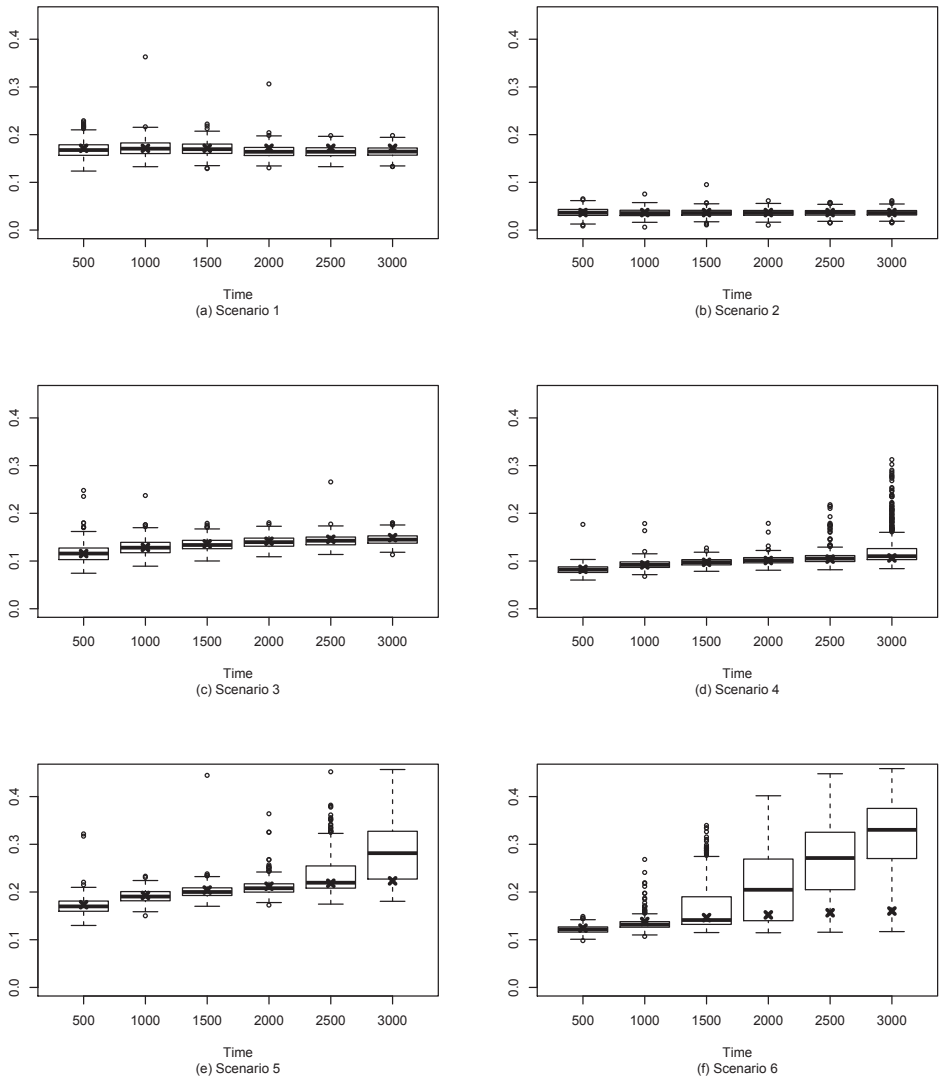

Figure 8: Results of the simulation for β_1: boxplots of the 500 estimated values obtained in each scenario, a cross represents the true value of β_1.

Appendix: Bootstrap algorithm

STEP 1:

- Take pilot bandwidth $g_j = \sqrt{2}h_j$ (see Remark 1).
- Obtain $\hat{S}_{g_j}(y|x_i), i = 1, \ldots, n$, the Beran estimator of $S(y|x_i)$ using bandwidth g_j.
- Obtain $(\hat{\beta}_{0,g_j}(y), \hat{\beta}_{1,g_j}(y))^t$ using the weighted least squares method described in Section 3.
- Obtain $\tilde{S}_{g_j}(y|x_i) = \phi^{-1}(\hat{\beta}_{0,g_j}(y) + \hat{\beta}_{1,g_j}(y)x_i)$.
- Obtain $\hat{G}(y)$ the Kaplan-Meier estimator of the censoring distribution (assumed independent for the covariate). Note that we could also use $\hat{G}_{g_j}(y|x_i), i = 1, \ldots, n$, the Beran estimator of the censoring distribution in cases where the censoring distribution does depend on the covariate.

Repeating this step for all eventtimes y provides a model-based estimated survival function $\tilde{S}_{g_j}(.|x_i)$ and an estimated censoring distribution $\hat{G}(.)$.

STEP 2:

- For $l = 1, \ldots, B$, generate a bootstrap sample $\{(Z^\star_{li}, \delta^\star_{li}, x_i\}, i = 1, \ldots, n$, as follows (see Remark 2):
  - $Y^\star_{li} \sim \tilde{S}_{g_j}(.|x_i)$
  - $C^\star_{li} \sim \hat{G}_{g_j}(.|x_i)$

- $Z^\star_{li} = min(Y^\star_{li}, C^\star_{li})$, $\delta^\star_{li} = 1_{\{Y^\star_i \leq C^\star_i\}}$

- For each bootstrap sample

  - Obtain $\hat{S}^\star_{lh_j}(y|x_i), i = 1, \ldots, n$, the Beran estimator of $S(y|x_i)$ using bandwidth h_j and data from the l^{th} bootstrap sample.

  - Obtain $\hat{\beta}^\star_{0,lh_j}(y), \hat{\beta}^\star_{1,lh_j}(y)$ using the weighted least squares method.

Remark: To draw random observations from a survival distribution, we use the idea that a survival function is actually a step function with successive jump $w_{(l)}$ such that $S(y_{(l-1)}) - S(y_{(l)}) = P(Y > y_{(l-1)}) - P(Y > y_{(l)}) = P(Y = y_{(l)}) = w_{(l)}$. So Y takes value $y_{(l)}$ with probability $w_{(l)}$ and therefore $Y^\star$ (generic notation for the $Y^\star_{li}$) is a random observation with discrete distribution $(y_{(l)}, w_{(l)}), l = 1, \ldots, r$. There is however a problem if the last observed time $y_{(r)}$ is censored as we will then not be able to define the last jump $w_{(r)}$. In such a case, we use the following solution. If the problem arises when drawing survival time from $\tilde{S}_{g_j}(.|x_i)$, we assign an arbitrary very large value (such that it will anyhow be censored at next step of the bootstrap algorithm) with probability $1 - \sum w_{(l)}$ and if it arises when generating censoring time from $\hat{G}(.|x_i)$, we assign as last value a maximum value (e.g. longest follow-up time in data) with probability $1 - \sum w_{(l)}$. To ensure that the survival distribution $\tilde{S}_{g_j}(.|x_i)$ is monotonic (and therefore the computed jump $w_{(l)}$ are positive), we use isotonic regression as mentioned in Section 5.