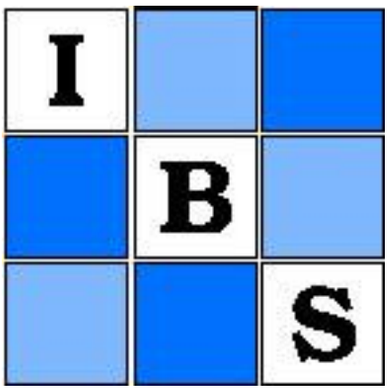

Piecewise Exponential Survival Trees with Time-Dependent Covariates

Author(s): Xin Huang, Shande Chen and Seng-jaw Soong

Source: *Biometrics*, Dec., 1998, Vol. 54, No. 4 (Dec., 1998), pp. 1420-1433

Published by: International Biometric Society

Stable URL: https://www.jstor.org/stable/2533668

REFERENCES
Linked references are available on JSTOR for this article:
https://www.jstor.org/stable/2533668?seq=1&cid=pdf-reference#references_tab_contents
You may need to log in to JSTOR to access the linked references.

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at https://about.jstor.org/terms

International Biometric Society is collaborating with JSTOR to digitize, preserve and extend access to *Biometrics*

This content downloaded from
210.117.171.92 on Tue, 27 Jul 2021 04:53:15 UTC
All use subject to https://about.jstor.org/terms

BIOMETRICS 54, 1420–1433
December 1998

Piecewise Exponential Survival Trees with Time-Dependent Covariates

Xin Huang,[1,*] Shande Chen,[2] and Seng-jaw Soong[3]

[1]Biometrics Department, Parke-Davis Pharmaceutical Research,
Ann Arbor, Michigan 48105, U.S.A.
[2]Department of Preventive Medicine, Rush-Presbyterian-St. Luke's Medical Center,
Chicago, Illinois 60612, U.S.A.
[3]Biostatistics Unit, Comprehensive Cancer Center,
University of Alabama at Birmingham, Birmingham, Alabama 35294, U.S.A.

SUMMARY

Survival trees methods are nonparametric alternatives to the semiparametric Cox regression in survival analysis. In this paper, a tree-based method for censored survival data with time-dependent covariates is proposed. The proposed method assumes a very general model for the hazard function and is fully nonparametric. The recursive partitioning algorithm uses the likelihood estimation procedure to grow trees under a piecewise exponential structure that handles time-dependent covariates in a parallel way to time-independent covariates. In general, the estimated hazard at a node gives the risk for a group of individuals during a specific time period. Both cross-validation and bootstrap resampling techniques are implemented in the tree selection procedure. The performance of the proposed survival trees method is shown to be good through simulation and application to real data.

1. Introduction

The Cox proportional hazards model (1972) and its extensions have been established as the major framework for regression analysis of survival data over two decades. The most important conceptual aspect of Cox's (1972) paper is that the hazard function is the basis of the regression model. The Cox method can easily handle time-dependent covariates, and such procedures are available in many statistical packages. Although the Cox proportional hazards model is a flexible tool for the study of covariate associations with survival time, it does not directly lead to models for prognostic groups. As in other regression models, the interactions in the Cox regression model are often modeled artificially. It is noted that the Cox regression model is sometimes used arbitrarily without proper model validation checks (Arjas, 1988) and that interpretation of the results from a model with time-dependent covariates is difficult (Andersen, 1991).

Recently, there has been growing interest in using tree-based methods in survival analysis because they are applicable to more general situations than classical regression approaches. In statistical research, there is a need to develop new approaches that perform better under fewer conditions and deal with interactions among variables in a more natural way. The demand for survival tree methods also comes from clinical studies in which the investigators are usually interested in grouping patients with differing prognoses.

Original tree-based methods introduced by Morgon and Sonquist (1963) were used in classification and regression. Advances in the practical and theoretical aspects of tree-based methods were developed by Breiman et al. (1984) in their monograph *Classification and Regression Trees*

* *Corresponding author's email address:* xin.huang@wl.com

Key words: Classification and regression trees; Survival analysis; Survival trees; Time-dependent covariates.

This content downloaded from
210.117.171.92 on Tue, 27 Jul 2021 04:53:15 UTC
All use subject to https://about.jstor.org/terms

(CART). Generally, tree-based methods recursively partition the covariate space into disjoint regions and the corresponding data into groups. For each node to be split, some measure of separation in the response distribution between the two daughter nodes is calculated. All possible splits for each of the covariates are evaluated, and the variable to be split and the split point are chosen to best separate the nodes. The same procedure is applied recursively to increase the number of nodes until each contains only a few subjects. The resulting model can be represented as a binary tree. After a large tree is grown, there are rules for recombining nodes and for readjusting the size of the tree.

Gordon and Olshen (1985) presented the first adaptation of CART to censored survival data, using distance measures between nearest continuous approximation of Kaplan–Meier curves. Davis and Anderson (1989) proposed a method based on the exponential log likelihood at nodes. Therneau, Grambsch, and Fleming (1990) proposed a method in which Martingale residuals were used directly in the CART regression algorithm with squared error loss. LeBlanc and Crowley (1992) extended the proportional hazards regression to tree-structured relative risk estimates for censored survival data with a one-step full likelihood estimation procedure. All four of these methods are based on measures of homogeneity within a node so that the application of CART is straightforward. Segal (1988) presented a totally nonparametric application using the Harrington–Fleming (1982) classes of two-sample rank statistics that based the partitioning on between-node separation instead of within-node homogeneity. Later, Bacchetti and Segal (1995) further extended this method to allow for truncation and time-dependent covariates. LeBlanc and Crowley (1993) developed a recursive partitioning procedure based on maximizing the dissimilarity in the survival distributions of patients between regions of the covariate space. These two methods are based on maximizing the dissimilarity in survival distribution between different regions of the covariate space. Ahn and Loh (1994) developed a tree-structured proportional hazards regression model that stratifies data according to selected covariate values and fits separate proportional hazards models to each stratum. Most of the existing survival trees methods are suitable only for dealing with time-independent covariates. Little work has been done on time-dependent covariates.

Time-dependent covariates often appear in survival analysis. Including such covariates leads to dynamic prognosis in which the relative risk can change from one time point to the next as the values of the covariates change. The study of time-dependent covariates in survival analysis has recently received considerable attention in both statistical and biomedical literature (Cox and Oakes, 1984; Andersen, 1991; Collett, 1994; White and Pocock, 1996). In this paper, we propose a model that accommodates time-dependent covariates into survival trees for censored survival data on the basis of piecewise exponential distributions. This method splits nodes through the interaction of the covariate values and time and establishes measures of improvement on the basis of piecewise exponential survival functions. The estimated hazard function at each branch node summarizes the risk of a group of subjects during each specific time period. This paper is organized as follows. Section 2 describes the procedure of the piecewise exponential survival trees. Section 3 presents several simulation studies on the proposed method. Section 4 exemplifies the method by reanalyzing the Stanford heart transplant data. Finally, Section 5 consists of some remarks and discussion.

2. Piecewise Exponential Survival Trees Method

We assume that survival data include the time to event and covariates that might be associated with time. Let $X = \min(T, C)$ be the observed time, where T is the true survival time and C is the true censoring time. Define the death indicator $\delta = 1$ if $X = T$ and $\delta = 0$ otherwise. Let $\mathbf{Z}$ be a vector of covariates, possibly with some time-dependent components. Then, an observation is denoted as a triple $(X, \delta, \mathbf{Z})$. As usual in survival analysis, we assume that T and C are independent conditional on $\mathbf{Z}$. For given $\mathbf{Z}$, the sample consists of a set of independent and identically distributed vectors $\{(x_i, \delta_i, \mathbf{z}_i) : i = 1, 2, \ldots, N\}$.

As with most existing survival trees, we work primarily on the hazard function. Instead of having a proportional assumption, we model the hazard function by a general form

$$\lambda(t) = \lambda(\mathbf{z}(t), t). \tag{1}$$

The proposed method approximates (1) by a piecewise constant function of t, where the jump points are chosen by the tree method detailed later in this section, and thus it approximates the survival experience of a subject by a piecewise exponential distribution. With such an approximation, it is not difficult to handle time-dependent covariates, and applying the CART principle is

This content downloaded from
210.117.171.92 on Tue, 27 Jul 2021 04:53:15 UTC
All use subject to https://about.jstor.org/terms

straightforward. Computationally, the argument t in $\lambda(\mathbf{z}, t)$ is treated as a time-dependent covariate in the piecewise exponential survival trees algorithm.

If all covariates are time independent, then (1) reduces to

$$\lambda(t) = \lambda(\mathbf{z}, t). \tag{2}$$

Existing survival trees methods (Gordon and Olshen, 1985; Segal, 1988; Davis and Anderson, 1989; Therneau et al., 1990; Leblanc and Crowley, 1992, 1993) approximate $\lambda(\mathbf{z}, t)$ in different functions for each of the homogeneous subject subgroups based on covariates $\mathbf{z}$, where only overall survival experience among subgroups is compared. With piecewise exponential survival trees, additional comparisons of the hazards in various time periods for the same subject are made. As a result of the convergence of the piecewise exponential approximation, piecewise exponential survival trees are asymptotically equivalent to the other nonparametric survival trees applying to (2). If argument t is excluded from (2), the proposed method produces exactly the same trees as the exponential survival trees method (Davis and Anderson, 1989).

In practice, the values of time-dependent covariates are usually measured at certain time points. Thus, $\mathbf{z}(t)$ is often treated as a step function. For computational reasons, we can work on an approximate model to (1),

$$\lambda(t) = \lambda(\mathbf{z}(t), a(t)), \tag{3}$$

where $a(t)$ is a nondecreasing step function of t. In (3), $a(t)$ plays the same role as t in (1). Replacing (1) with (3) can reduce a great amount of computing time and resources while maintaining reasonably good results.

With the hazard model given in (1) or (3), we can present details of the piecewise exponential survival trees method. Although many types of partitions could be considered, we consider only the splits on a single variable at a time, which can be easily generalized to combinations of covariates. Because the argument t is treated as a time-dependent covariate in the algorithm, as illustration we consider only a single time-dependent covariate $Z(t)$ in the hazard function. Thus, for the subject i $(i = 1, 2, \ldots, N)$, the risk at a certain time is a function solely of the covariate value at that time. The piecewise exponential survival trees method approximates the relationship between the hazard and the value of the covariate by a step function that is illustrated in Figure 1. Upper-left and lower-right graphs show the changes of the hazard and the time-dependent covariate with time, respectively. Assume that both the hazard and the time-dependent covariate are continuous functions of time that are represented by solid smooth curves. The upper-right graph demonstrates the hazard as a continuous function of the value of the covariate and corresponding step approximations. With this step approximation, the hazard is approximated by a step function of t, shown in solid lines in the upper-left graph. Dashed lines show the connection between the hazard and the covariate in each time interval. In general, the algorithm approximates distribution of T_i by a piecewise exponential (PE) distribution of k pieces, which has a density function

$$f_i(t) = \begin{cases} \lambda_{i_1} \exp(-\lambda_{i_1} t), & 0 = t_{i_0} < t \le t_{i_1}, \\ \lambda_{i_2} \exp(t_{i_1}(\lambda_{i_2} - \lambda_{i_1}) - \lambda_{i_2} t), & t_{i_1} < t \le t_{i_2}, \\ \vdots & \vdots \\ \lambda_{i_k} \exp\left[\sum_{j=1}^{k-1} t_{i_j}(\lambda_{i_{j+1}} - \lambda_{i_j}) - \lambda_{i_k} t\right], & t_{i_{k-1}} < t \le t_{i_k}, \end{cases}$$

where $0 = t_{i_0} < t_{i_1} < \cdots < t_{i_k} = \infty$ and $\lambda_{i_1}, \lambda_{i_2}, \ldots, \lambda_{i_k}$ are positive. For mathematical convenience, some t_{i_j} $(1 < j < k)$ can be defined as ∞ so that $f_i(t)$ can have less than k pieces. It is well known that piecewise exponential distribution can provide good approximations for various survival models with reasonable accuracy for a variety of shapes for the hazard.

With such a survival trees model, the application of the CART principle is straightforward. For simplicity, we illustrate the case with only one time-dependent covariate $Z(t)$, which is assumed to be nondecreasing in time for each subject. It can be handled analogously if $Z(t)$ is nonincreasing or $Z(t)$ is nonmonotone. At the initial splitting with a split point S of $Z(t)$, subject i can belong to one of three categories: (1) subject i with $z(t) \le S$ at all times goes to the left daughter node, (2) subject i with $z(t) > S$ at all times goes to the right daughter node, and (3) subject i with $z(t) \le S$ when $0 < t \le t_{i,s}$ contributes to the left daughter node, and $z(t) > S$ when $t_{i,s} < t \le x_i$ goes to the right daughter node. The third category appears only when a time-dependent covariate exists, which indicates that subject i's survival experiences are different in two nonoverlapping time

This content downloaded from
210.117.171.92 on Tue, 27 Jul 2021 04:53:15 UTC
All use subject to https://about.jstor.org/terms

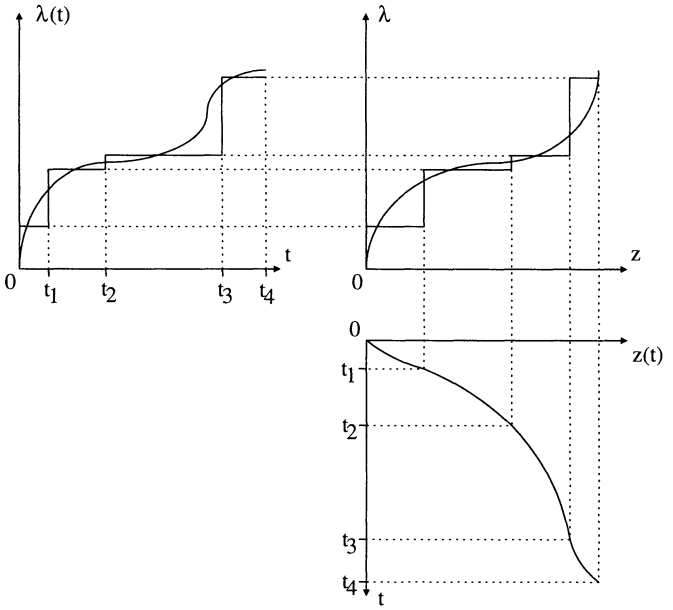

Figure 1. Relationship between a hazard function and a time-dependent covariate. Upper-left and lower-right graphs show the changes of the hazard and the time-dependent covariate with time, respectively. Assume that the hazard and the time-dependent covariate are both continuous functions of time that are represented by solid smooth curves. The upper-right graph demonstrates the hazard as a continuous function of the value of the covariate and corresponding step approximations, which also are shown in solid lines. With this step approximation, the hazard is approximated by a step function of T shown in solid lines in the upper-left graph. Dashed lines show the connection between the hazard and the covariate in each time interval.

intervals. The piecewise exponential survival trees model we propose here requires that subjects i be considered part of the left daughter node when $0 < t \leq t_{i,s}$ and part of the right daughter node when $t_{i,s} < t \leq x_i$. This is further illustrated with Stanford heart transplant data in Section 4.

The new split point is selected on the basis of the piecewise exponential survival. Suppose that the tree is grown to the $(k-1)$th stage, and thus there are k nodes. Let $S_{(1)} < S_{(2)} < \cdots < S_{(k-1)}$ be the ordered values of split points $S_1, S_2, \ldots, S_{k-1}$. Let $\Gamma_{(j)}$, $j = 1, 2, \ldots, k$, denote the nodes given by $z(t) \leq S_{(1)}$, $S_{(j-1)} < z(t) \leq S_{(j)}$ $(1 < j < k)$, or $z(t) > S_{(k-1)}$. Let S_k be a further split point for time-dependent covariate $z(t)$, and assume that $S_{(j-1)} < S_k < S_{(j)}$ $(1 < j < k)$. Thus, there is a new split at node $\Gamma_{(j)}$. The left daughter consists of subjects in $\Gamma_{(j)}$ subjected to $S_{(j-1)} < z(t) \leq S_k$, and the right daughter consists of subjects in $\Gamma_{(j)}$ subjected to $S_k < z(t) \leq S_{(j)}$.

In general, before splitting, the log-likelihood function of this node is

$$l_M(\lambda_k) = \sum_{i \in I_1} [\delta_i \log(\lambda_k) - \lambda_k x_i] - \sum_{i \in I_2} \lambda_k t_{i,(j)} + \sum_{i \in I_3} \left[\delta_i \log(\lambda_k) - \lambda_k (x_i - t_{i,(j-1)})\right] - \sum_{i \in I_4} \sum \lambda_k (t_{i,(j)} - t_{i,(j-1)}),$$

where $I_1 = \{i \mid S_{(j-1)} < z(t) \leq S_{(j)},\ 0 < t \leq x_i\}$, $I_2 = \{i \mid S_{(j-1)} < z(t) \leq S_{(j)},\ 0 < t \leq t_{i,(j)};\ z(x_i) > S_{(j)}\}$, $I_3 = \{i \mid z(t) \leq S_{(j-1)},\ 0 < t \leq t_{i,(j-1)};\ S_{(j-1)} < z(t) \leq S_{(j)},\ t_{i,(j-1)} < t \leq x_i\}$, and $I_4 = \{i \mid z(t) \leq S_{(j-1)},\ 0 < t \leq t_{i,(j-1)};\ S_{(j-1)} < z(t) \leq S_{(j)},\ t_{i,(j-1)} < t \leq t_{i,(j)};\ z(x_i) > S_{(j)}\}$; and $z(t_{i,(j-1)}) = S_{(j-1)}$ and $z(t_{i,(j)}) = S_{(j)}$.

For a simple case, such as $j = 3$ and $k = 5$, we have a diagram showing the graphs of $Z(t)$ versus t for four hypothetical subjects, one belonging to each of I_1, I_2, I_3, and I_4. Gray regions show how each I_i $(i = 1, 2, 3, 4)$ is defined. From Figure 2a, we can see that subject i belonging to I_1 has $S_{(2)} < z(t) \leq S_{(3)}$ at all times $(0 < t \leq x_i)$. In Figure 2b, on the other hand, subject i belonging to I_2 has $S_{(2)} < z(t) \leq S_{(3)}$ when $0 < t \leq t_{i,(3)}$ and $S_{(3)} = z(t_{i,(3)}) < z(t) \leq z(x_i)$ when $t_{i,(3)} < t \leq x_i$. Contrasting with subjects in I_2, in Figure 2c, subject i belonging to I_3 has

This content downloaded from
210.117.171.92 on Tue, 27 Jul 2021 04:53:15 UTC
All use subject to https://about.jstor.org/terms

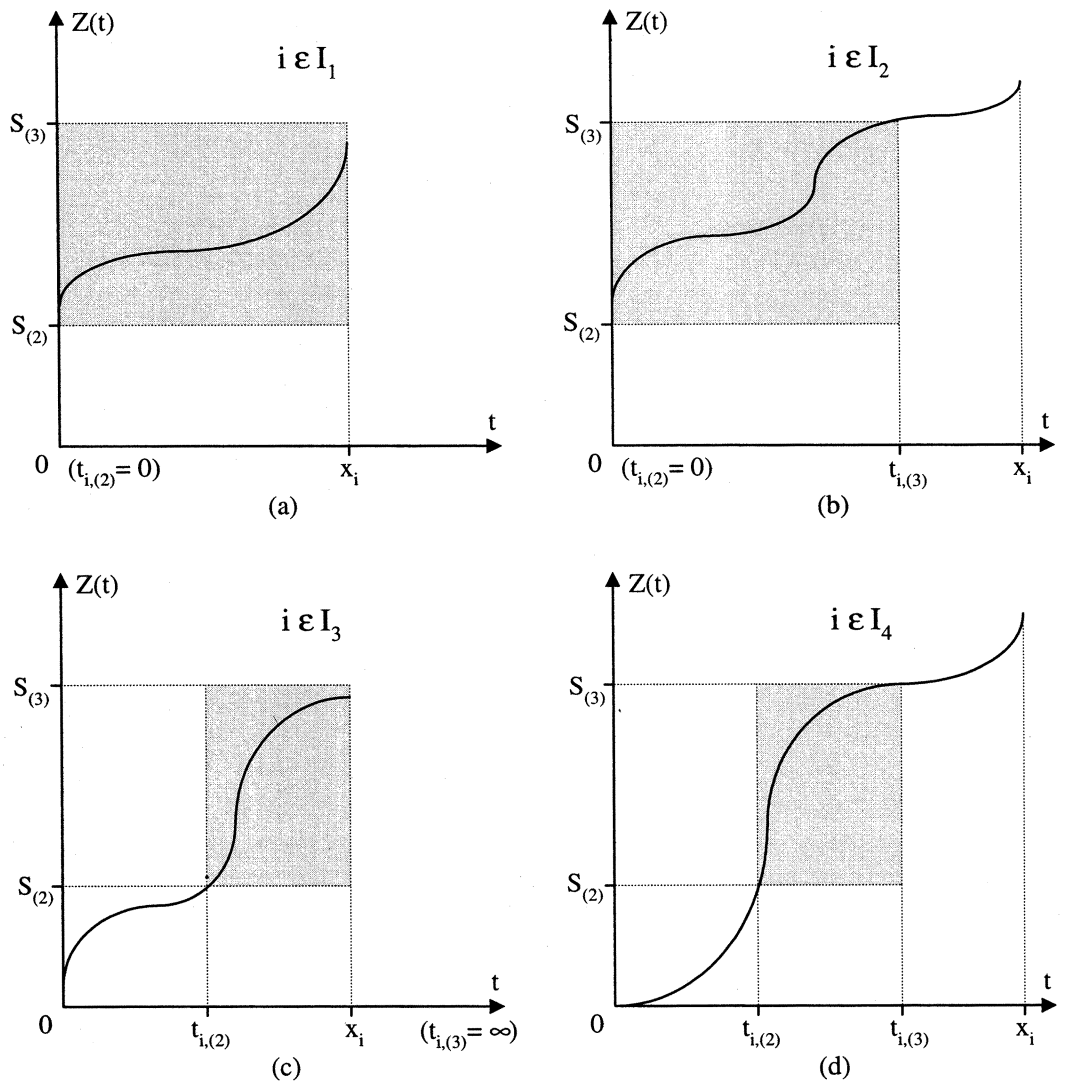

Figure 2. Graphs of $Z(t)$ versus time for four hypothetical subjects, one from each of I_1, I_2, I_3, and I_4. Gray regions show how each I_i, $(i = 1, 2, 3, 4)$ is defined. **a.** Subject i belonging to I_1 has $S_{(2)} < z(t) \leq S_{(3)}$ at all times $(0 < t \leq x_i)$. **b.** Subject i belonging to I_2 has $S_{(2)} < z(t) \leq S_{(3)}$ when $0 < t \leq t_{i,(3)}$ and $S_{(3)} = z(t_{i,(3)}) < z(t) \leq z(x_i)$ when $t_{i,(3)} < t \leq x_i$. **c.** Subject i belonging to I_3 has $z(t) \leq S_{(2)} = z(t_{i,(2)})$ when $0 < t \leq t_{i,(2)}$ and $S_{(2)} < z(t) \leq S_{(3)}$ when $t_{i,(2)} < t \leq x_i$. **d.** Subject i belonging to I_4 has $z(t) \leq S_{(2)} = z(t_{i,(2)})$ when $0 < t \leq t_{i,(2)}$, $S_{(2)} < z(t) \leq S_{(3)}$ when $t_{i,(2)} < t \leq t_{i,(3)}$, and $S_{(3)} = z(t_{i,(3)}) < z(t) \leq z(x_i)$ when $t_{i,(3)} < t \leq x_i$.

$z(t) \leq S_{(2)} = z(t_{i,(2)})$ when $0 < t \leq t_{i,(2)}$ and $S_{(2)} < z(t) \leq S_{(3)}$ when $t_{i,(2)} < t \leq x_i$. Finally, in Figure 2d, subject i belonging to I_4 has not only $z(t) \leq S_{(2)} = z(t_{i,(2)})$ when $0 < t \leq t_{i,(2)}$ but also $S_{(2)} < z(t) \leq S_{(3)}$ when $t_{i,(2)} < t \leq t_{i,(3)}$ and $S_{(3)} = z(t_{i,(3)}) < z(t) \leq z(x_i)$ when $t_{i,(3)} < t \leq x_i$.

The maximum likelihood estimate of λ_k is

$$\hat{\lambda}_k = \frac{\sum_{i \in I_1} \delta_i + \sum_{i \in I_3} \delta_i}{\sum_{i \in I_1} x_i + \sum_{i \in I_2} t_{i,(j)} + \sum_{i \in I_3} \left(x_i - t_{i,(j-1)}\right) + \sum_{i \in I_4} \left(t_{i,(j)} - t_{i,(j-1)}\right)}. \tag{4}$$

This content downloaded from
210.117.171.92 on Tue, 27 Jul 2021 04:53:15 UTC
All use subject to https://about.jstor.org/terms

The maximum likelihood after the splitting is similar, and the estimates of λ_{kl} and λ_{kr} for the left and right daughters of node $\Gamma_{(j)}$ are

$$\hat{\lambda}_{kl} = \frac{\sum_{i \in L_1} \delta_i + \sum_{i \in L_3} \delta_i}{\sum_{i \in L_1} x_i + \sum_{i \in L_2} t_{i,k} + \sum_{i \in L_3} \left(x_i - t_{i,(j-1)}\right) + \sum_{i \in L_4} \left(t_{i,k} - t_{i,(j-1)}\right)}$$

and

$$\hat{\lambda}_{kr} = \frac{\sum_{i \in R_1} \delta_i + \sum_{i \in R_3} \delta_i}{\sum i \in R_1 x_i + \sum_{i \in R_2} t_{i,(j)} + \sum_{i \in R_3} \left(x_i - t_{i,k}\right) + \sum_{i \in R_4} \left(t_{i,(j)} - t_{i,k}\right)},$$

where L_j and R_j $(j = 1, 2, 3, 4)$ are the sets corresponding to I_j $(j = 1, 2, 3, 4)$ for the left and right daughters of $\Gamma_{(j)}$.

There are two special cases. First, if the node $\Gamma_{(1)}$ is split, then I_3, I_4, L_3, and L_4 are empty. Second, if the node $\Gamma_{(k)}$ is split, then I_2, I_4, R_2, and R_4 are empty. For these two special cases, the likelihood functions and the parameter estimates for daughter nodes are slightly simpler.

The loss functions are defined as $R_M(\lambda) = -l_M(\lambda)$, $R_L(\lambda) = -l_L(\lambda)$, and $R_R(\lambda) = -l_R(\lambda)$. Thus, the improvement of the new split is

$$R\left(\Gamma_{(j)}; S_k\right) = R_M\left(\hat{\lambda}_k\right) - \left[R_L\left(\hat{\lambda}_{kl}\right) + R_R\left(\hat{\lambda}_{kr}\right)\right].$$

The algorithm evaluates all the possible values for S_k and selects the one that maximizes this improvement as the new splitting point and grows the tree to the kth stage.

The binary splitting continues until the number of subjects in each node reaches a given minimum node size. If a time-dependent covariate is chosen to split a node, it is possible that a subject contributes to both daughter nodes but is associated with two disjoint time periods corresponding to the value of this time-dependent covariate. Thus, the estimated hazards at each daughter node are the assessment of the risk of this subject during each of the time periods. As the tree grows, the relative risk functions might have many pieces.

After a complete tree (possibly quite large and overfitted) is grown, the prune procedure is essentially the same as in CART. Let $\tilde{T}$ be the set of terminal nodes for a tree T, and let $\left\|\tilde{T}\right\|$ denote the number of terminal nodes. Following Breiman et al. (1984), the cost complexity of a piecewise exponential tree is defined to be

$$R_\alpha(T) = \sum_{\Gamma \in \tilde{T}} R(\Gamma) + \alpha \left\|\tilde{T}\right\|$$

for a nonnegative complexity parameter α, where the impurity $R(\Gamma)$ is defined as

$$R(\Gamma) = -l(\hat{\lambda}_\Gamma) = \sum_{i \in I_\Gamma} (1 - \log \hat{\lambda}_\Gamma)\delta_i,$$

where $\hat{\lambda}_\Gamma$ is estimated hazard as described in (4). CART provides an efficient algorithm for obtaining the optimal sequence of subtrees, which has the minimum cost-complexity measures.

As with CART, a nested sequence of subtrees is defined by minimal cost-complexity pruning. Either cross-validation or bootstrap resampling (Efron, 1982; Efron and Tibshirani, 1993) can be used to make honest estimates of the loss associated with each tree in the sequence, and the final tree is selected on the basis of these estimates. Generally, the two methods are asymptotically the same, although they can behave quite differently in different situations. Early simulation experiments (Efron, 1982; Efron and Tibshirani, 1993) showed that cross-validation was roughly unbiased but could show large variability, whereas the bootstrap method had lower variability but could be severely biased downward. From our limited experience, either method works fine for the piecewise exponential trees.

3. Simulation Studies

To investigate the performance of the piecewise exponential survival trees, simulation studies were conducted. In this section, procedures and results of four simulation experiments are reported.

This content downloaded from
210.117.171.92 on Tue, 27 Jul 2021 04:53:15 UTC
All use subject to https://about.jstor.org/terms

Because most of the existing survival trees methods are suitable for dealing with time-independent covariates, comparisons are made only between the proposed method and the Cox (1972) regression method. Considering the computing-intensive nature of the tree-based methods, we conducted only relatively small simulations.

Each of the four simulations was completed with different perspectives. Simulation I was conducted to investigate the behavior of the proposed method in a continuously proportional hazards setting. Simulations II and III were designed to examine the performance of the proposed method when the hazard is either a monotone or a nonmonotone function of the covariate. Simulation IV was used to study the situation in which both time-dependent and time-independent covariates exist.

We used nondecreasing step functions for the structure of our time-dependent covariates and limited only a few possible jump points in simulations. Obviously, the model can handle continuous time-dependent covariates, but it would be difficult to summarize the simulations were a continuous variable used because of many possible cut points. In simulation I, failure times were generated by Weibull distributions associated with a time-independent covariate. In simulations II and III, survival times were generated by piecewise exponential distributions associated with a single time-dependent covariate. In simulation IV, the piecewise exponential survival was associated with a time-dependent and a time-independent covariate. Censoring times were simulated from the uniform $(0, \gamma)$ distribution, with γ chosen to have a censoring rate between 15% and 20%. In each simulation, 100 random samples were generated. The minimum terminal node size permitted for splitting was 20 observations. The selection of right-sized pruned subtrees was based on estimates of the prediction errors from minimizing the 10-fold cross-validation.

The piecewise exponential survival trees method estimates primarily the hazards. In contrast, the Cox regression model estimates the relative risks with respect to an unspecified baseline hazard. Thus, we can choose the hazard function of one group as the baseline hazard and obtain all hazard ratio estimates from both methods for comparison. Procedure *phreg* of SAS was used in simulations for the Cox model.

We would like to point out that tree-based models are invariant to monotone transformations of covariates so that the precise form in which they appear in the analysis is irrelevant. On the other hand, in most applications, this is not true for ordinary regression models. To obtain correct estimates, dummy variables constructed by complicated programming statements were used in the Cox model throughout simulations.

Four hundred random subjects were generated by Weibull distributions with a constant shape parameter ($\kappa = 1.5$) and three different values of the scale parameter ($\rho = 1, 2, 3$) in simulation I. The allocation ratios were 1:1:1. The scale parameter was treated as a time-independent covariate. The corresponding proportional hazard functions were determined by $\lambda(t) = 1.5\rho^{3/2}t^{1/2}, \rho = 1, 2, 3$. This was exactly a proportional hazards model. For computational reasons, we used a step function (as a time-dependent covariate) to approximate the time.

In simulation I, hazard functions given by $\rho = 1$, 2, and 3 were denoted as λ_1, λ_2, and λ_3. Two given hazard ratios were $\lambda_1/\lambda_2 = 0.3536$ and $\lambda_1/\lambda_3 = 0.1925$. The means of the estimated hazard ratios from the Cox model were 0.3526 and 0.1997 with standard errors of 0.0546 and 0.0374 and mean square errors of 0.0030 and 0.0014 for those hazard ratio estimates, respectively. The mean square error was defined as the average sum square of the difference between the given hazard ratio and its estimate in each of the 100 repetitions. For simplicity, we used a step function with only one jump point to approximate time for the proposed model. The jump point was set at 1, which resulted in a roughly equal number of subjects in two time periods $[0, 1)$ and $[1, \infty)$. The means of the estimated hazard ratios from the proposed model at the first time period were 0.3876 and 0.2310 with standard errors of 0.0740 and 0.0566 and mean square errors of 0.0067 and 0.0047, respectively. The means of the estimated hazard ratios from the proposed model at the second time period were 0.4065 and 0.2516 with standard errors of 0.0810 and 0.0515 and mean square errors of 0.0093 and 0.0061, respectively. The results indicate that the performance of the proposed method is good in this continuous proportional hazards setting. The piecewise exponential survival trees model is able to cope with continuous hazard functions by only a rough approximation of two-piece exponential distribution. It will perform better when more pieces for piecewise exponential distribution are allowed.

In each repetition of simulation II, 400 subjects were generated by piecewise exponential distributions. The time-dependent covariate $z(t)$ was modeled by four step functions, each having three possible common jump points. The underlying hazards were a monotone nondecreasing function of the time-dependent covariate. We denoted λ_{p_q}, $p = 1, 2, 3, 4$ and $q = 1, 2, 3, 4$, as the hazard

This content downloaded from
210.117.171.92 on Tue, 27 Jul 2021 04:53:15 UTC
All use subject to https://about.jstor.org/terms

associated with pth step (covariate) function at the qth time period of $[0,3)$, $[3,6)$, $[6,9)$, and $[9,\infty)$. The first step function for $z(t)$ was set as 1, 1, 1, and 1 at each of the four time periods. A corresponding piecewise hazard was chosen as $\lambda_{1_1} = \lambda_{1_2} = \lambda_{1_3} = \lambda_{1_4} = 0.05$, which served as our baseline hazard. A piecewise hazard was chosen as $\lambda_{2_1} = 0.025$, $\lambda_{2_2} = 0.05$, $\lambda_{2_3} = 0.10$, and $\lambda_{2_4} = 0.15$ to correspond to the second step function for $z(t)$, which took values 0, 1, 2, and 3 at each of the four time periods. Similarly, piecewise hazards were chosen as $\lambda_{3_1} = \lambda_{3_2} = 0.025$, $\lambda_{3_3} = 0.10$, $\lambda_{3_4} = 0.20$ and $\lambda_{4_1} = 0.05$, $\lambda_{4_2} = 0.10$, $\lambda_{4_3} = 0.15$, $\lambda_{4_4} = 0.30$ to correspond to the third step function and the fourth step function for $z(t)$, which took values 0, 0, 2, and 4 and 1, 2, 3, and 5 at each of the four time periods, respectively. The number of subjects associated with each of the four step functions was approximately equal.

Survival trees with six distinct terminal nodes were grown by the piecewise exponential survival trees model in simulation II. They provided estimates of six values, 0.025, 0.05, 0.10, 0.15, 0.20, and 0.30, for hazards. It appeared that λ_{2_1}, λ_{3_1}, and λ_{3_2}; λ_{1_1}, λ_{1_2}, λ_{1_3}, λ_{1_4}, λ_{2_2}, and λ_{4_1}; λ_{2_3}, λ_{3_3}, and λ_{4_2}; λ_{2_4} and λ_{4_3} had identical estimates, respectively. With one set of estimated piecewise hazards as the baseline, we compared the simulation results between the proposed method and Cox model. The summary statistics of estimated hazard ratios from piecewise exponential survival trees and Cox regression model are presented in Table 1. For a given hazard ratio, the mean of estimated hazard ratios (Mean), the standard error of the mean (SE), and the mean square error (MSE) is listed. Some hazard ratio estimates are identical because estimates of both their numerators and their denominators are identical as well. For example, in Table 1 $\lambda_{2_1}/\lambda_{1_1}$, $\lambda_{3_1}/\lambda_{1_1}$, and $\lambda_{3_2}/\lambda_{1_2}$ all have the same hazard ratio estimates 0.5126 from survival trees and 0.5198 from the Cox model, respectively. There is a special case in which both the numerator and the denominator of a hazard ratio are identical. If this happens, the hazard ratio estimates are listed as 1.0 in the table. Table 1 shows that both the piecewise exponential survival trees method and the Cox regression performed well.

In simulation III, nonmonotone relations between the covariate and the hazards were implemented. Four different piecewise exponential distributions, each with three possible common jump points, were used to generate 400 subjects of each repetitions. The underlying hazards were defined in a similar way as simulation II, and hazards were also represented as λ_{p_q}, $p = 1,2,3,4$ and $q = 1,2,3,4$. The first step function of for $z(t)$ was set as 1, 1, 2, and 2 at each of the four time periods $[0,3)$, $[3,6)$, $[6,9)$, and $[9,\infty)$. A corresponding piecewise hazard was chosen as $\lambda_{1_1} = \lambda_{1_2} = 0.05$ and $\lambda_{1_3} = \lambda_{1_4} = 0.10$, which served as our baseline hazard. A piecewise hazard was chosen as $\lambda_{2_1} = 0.05$, $\lambda_{2_2} = 0.10$, $\lambda_{2_3} = 0.30$, and $\lambda_{2_4} = 0.20$ to correspond to the second

Table 1

Summary statistics of estimated hazard ratios from piecewise exponential survival (PES) trees and Cox regression model for simulation II. Means of estimated hazard ratios, standard errors of the means (SE), and mean square errors (MSE) are listed for both methods.

		Estimate					
		PES trees			Cox model		
Hazard ratio	Given value	Mean	SE	MSE	Mean	SE	MSE
$\lambda_{2_1}/\lambda_{1_1}$	0.5	0.5126	0.0125	0.0157	0.5198	0.0136	0.0183
$\lambda_{2_2}/\lambda_{1_2}$	1.0	1.0[a]			1.0		
$\lambda_{2_3}/\lambda_{1_3}$	2.0	2.0389	0.0397	0.1574	2.0925	0.0449	0.2077
$\lambda_{2_4}/\lambda_{1_4}$	3.0	3.0992	0.0526	0.2841	3.1785	0.0764	0.6102
$\lambda_{3_1}/\lambda_{1_1}$	0.5	0.5126	0.0125	0.0157	0.5198	0.0136	0.0183
$\lambda_{3_2}/\lambda_{1_2}$	0.5	0.5126	0.0125	0.0157	0.5198	0.0136	0.0183
$\lambda_{3_3}/\lambda_{1_3}$	2.0	2.0389	0.0397	0.1574	2.0925	0.0449	0.2077
$\lambda_{3_4}/\lambda_{1_4}$	4.0	3.9859	0.0640	0.4060	4.1201	0.0983	0.9717
$\lambda_{4_1}/\lambda_{1_1}$	1.0	1.0[a]			1.0		
$\lambda_{4_2}/\lambda_{1_2}$	2.0	2.0389	0.0397	0.1574	2.0925	0.0449	0.2077
$\lambda_{4_3}/\lambda_{1_3}$	3.0	3.0992	0.0526	0.2841	3.1785	0.0764	0.6102
$\lambda_{4_4}/\lambda_{1_4}$	6.0	6.1052	0.1091	1.1896	6.2169	0.1519	2.3311

Baseline hazards $\lambda_{1_1} = \lambda_{1_2} = \lambda_{1_3} = \lambda_{1_4} = 0.05$.

Some hazard ratio estimates are identical because estimates of both their numerator and their denominators are identical.

[a] The estimates of λ_{1_1}, λ_{1_2}, λ_{2_2}, and λ_{4_1} are identical.

This content downloaded from
210.117.171.92 on Tue, 27 Jul 2021 04:53:15 UTC
All use subject to https://about.jstor.org/terms

Table 2

Summary statistics of estimated hazard ratios from piecewise exponential survival (PES) trees and Cox regression model for simulation III. Means of estimated hazard ratios, standard errors of the means (SE), and mean square errors (MSE) are listed for both methods.

		Estimate					
		PES trees			Cox model		
Hazard ratio	Given value	Mean	SE	MSE	Mean	SE	MSE
$\lambda_{2_1}/\lambda_{1_1}$	1.0	1.0[a]			1.0		
$\lambda_{2_2}/\lambda_{1_2}$	2.0	1.9844	0.0296	0.0869	1.9835	0.0355	0.5133
$\lambda_{2_3}/\lambda_{1_3}$	3.0	3.0727	0.0406	0.1685	3.1737	0.0699	0.1253
$\lambda_{2_4}/\lambda_{1_4}$	2.0	2.0086	0.0330	0.1081	2.0593	0.0413	2.0218
$\lambda_{3_1}/\lambda_{1_1}$	1.0	1.0[a]			1.0[a]		
$\lambda_{3_2}/\lambda_{1_2}$	6.0	6.0555	0.0989	0.9710	6.1643	0.1420	0.1721
$\lambda_{3_3}/\lambda_{1_3}$	2.0	1.9844	0.0296	0.0869	1.9835	0.0355	0.5133
$\lambda_{3_4}/\lambda_{1_4}$	2.0	1.9844	0.0296	0.0869	1.9835	0.0355	0.5133
$\lambda_{4_1}/\lambda_{1_1}$	2.0	2.0086	0.0330	0.1081	2.0593	0.0413	2.0218
$\lambda_{4_2}/\lambda_{1_2}$	4.0	3.9405	0.0618	0.3811	4.0012	0.0848	0.1425
$\lambda_{4_3}/\lambda_{1_3}$	1.5	1.5643	0.0357	0.1304	1.5888	0.0369	0.7127
$\lambda_{4_4}/\lambda_{1_4}$	1.5	1.5643	0.0357	0.1304	1.5888	0.0369	0.7127

Baseline hazards $\lambda_{1_1} = \lambda_{1_2} = 0.05$, $\lambda_{1_3} = \lambda_{1_4} = 0.10$.

Some hazard ratio estimates are identical because both their numerator and their denominators are identical.

[a] The estimates of λ_{1_1}, λ_{2_1}, and λ_{3_1} are identical.

step function for $z(t)$, which took values 1, 2, 4, and 6 at each of the four time periods, respectively. Similarly, piecewise hazards were chosen as $\lambda_{3_1} = 0.05$, $\lambda_{3_2} = 0.30$, $\lambda_{3_3} = \lambda_{3_4} = 0.20$ and $\lambda_{4_1} = 0.10$, $\lambda_{4_2} = 0.20$, $\lambda_{4_3} = \lambda_{4_4} = 0.15$ to correspond to the third step function and the fourth step function for $z(t)$, which took values 1, 4, 6, and 6 and 2, 3, 5, and 5 at each of the four time periods, respectively. It appeared that the underlying hazards increased early, then decreased with the corresponding monotone increasing covariate.

The piecewise exponential survival trees model also grew survival trees with six distinct terminal nodes in simulation III. They provided estimates of six values, 0.05, 0.10, 0.30, 0.15, and 0.20, for hazards. It turned out that λ_{1_1}, λ_{1_2}, λ_{2_1}, and λ_{3_1}; λ_{1_3}, λ_{1_4}, λ_{2_2}, and λ_{4_1}; λ_{2_3}, and λ_{3_2}; λ_{4_3}, and λ_{4_4}; λ_{2_4} λ_{3_3} and λ_{3_4} had identical estimates, respectively. The summary statistics of estimated hazard ratios from piecewise exponential survival trees and the Cox regression model are shown in Table 2. For similar reasons as mentioned for simulation II, some of the hazard ratio estimates are identical in the table. As shown in Table 2, the piecewise exponential survival trees performed well. The Cox regression performed nearly as well as the proposed method.

In simulation IV, samples of $n = 450$ subjects whose survival times depended on a time-dependent and a time-independent covariate were generated. The time-dependent covariate $z(t)$ was modeled by three step functions, each with two possible common jump points. The time-independent covariate took three values. Fifty piecewise exponential survival times were generated for each of the nine combination of the two covariates. The underlying hazards were monotone increasing with both time-dependent and time-independent covariates. We denoted λ_{l,p_q}, $l = 1, 2, 3$, $p = 1, 2, 3$, and $q = 1, 2, 3$, as hazards associated with the lth value of the time-independent covariate, i.e., the pth step (time-dependent covariate) function at the qth time period of $[0, 3)$, $[3, 6)$, and $[6, \infty)$. For the first value of the time-independent covariate, piecewise hazards were chosen as $\lambda_{1,1_1} = \lambda_{1,1_2} = 0.02$, and $\lambda_{1,1_3} = 0.10$; $\lambda_{1,2_1} = 0.02$, $\lambda_{1,2_2} = 0.10$, and $\lambda_{1,2_3} = 0.30$; $\lambda_{1,3_1} = 0.10$, and $\lambda_{1,3_2} = \lambda_{1,3_3} = 0.30$ to correspond to the step functions for $z(t)$, which took values 1, 1, and 2; 1, 2, and 3; 2, 3, and 3 at each of the three time periods, respectively. The same pattern but different hazards were chosen for the other values of the time-independent covariate. The piecewise hazard $\lambda_{2,1_1} = \lambda_{2,1_2} = 0.10$, and $\lambda_{2,1_3} = 0.30$ was picked as the baseline hazard.

We obtained eight of the nine expected terminal nodes that corresponded to three values of the time-independent and three values of the time-dependent covariates in simulation IV. One terminal node could not be obtained because of an insufficient number of subjects to split further. Both the piecewise exponential survival trees and the Cox model give fairly good estimates. The summary statistics of estimated hazard ratios are presented in Table 3; the Cox model performed slightly

This content downloaded from
210.117.171.92 on Tue, 27 Jul 2021 04:53:15 UTC
All use subject to https://about.jstor.org/terms

Table 3

Summary statistics of estimated hazard ratios from piecewise exponential survival (PES) trees and Cox regression model for simulation IV. Means of estimated hazard ratios, standard errors of the means (SE), and mean square errors (MSE) are listed for both methods.

		Estimate					
		PES trees			Cox model		
Hazard ratio	Given value	Mean	SE	MSE	Mean	SE	MSE
$\lambda_{1,1_1}/\lambda_{2,1_1}$	0.2	0.2065	0.0109	0.0465	0.2082	0.1260	0.0158
$\lambda_{1,1_2}/\lambda_{2,1_2}$	0.2	0.2065	0.0109	0.0465	0.2082	0.1260	0.0158
$\lambda_{1,1_3}/\lambda_{2,1_3}$	0.333	0.3480	0.0066	0.0045	0.3449	0.0657	0.0044
$\lambda_{1,2_1}/\lambda_{2,1_1}$	0.2	0.2065	0.0109	0.0465	0.2082	0.1260	0.0158
$\lambda_{1,2_2}/\lambda_{2,1_2}$	1.0	1.0546	0.0322	0.1058	1.0212	0.1686	0.0286
$\lambda_{1,2_3}/\lambda_{2,1_3}$	1.0	1.0528	0.0193	0.0397	1.0212	0.1686	0.0286
$\lambda_{1,3_1}/\lambda_{2,1_1}$	1.0	1.0546	0.0322	0.1058	1.0212	0.1686	0.0286
$\lambda_{1,3_2}/\lambda_{2,1_2}$	3.0	3.1675	0.0821	0.6961	3.0931	0.8117	5.0331
$\lambda_{1,3_3}/\lambda_{2,1_3}$	1.0	1.0528	0.0193	0.0397	1.0212	0.1686	0.0286
$\lambda_{2,2_1}/\lambda_{2,1_1}$	1.0	1.0[a]			1.0		
$\lambda_{2,2_2}/\lambda_{2,1_2}$	3.0	3.0662	0.0837	0.6984	3.0048	0.4544	0.2044
$\lambda_{2,2_3}/\lambda_{2,1_3}$	1.667	1.8273	0.0470	0.2441	1.7084	0.2932	0.0869
$\lambda_{2,3_1}/\lambda_{2,1_1}$	3.0	3.0662	0.0837	0.6984	3.0048	0.4544	0.2044
$\lambda_{2,3_2}/\lambda_{2,1_2}$	5.0	5.4651	0.1574	2.6701	5.1427	1.2363	1.5335
$\lambda_{2,3_3}/\lambda_{2,1_3}$	1.667	1.8273	0.0470	0.2441	1.7084	0.2932	0.0869
$\lambda_{3,1_1}/\lambda_{2,1_1}$	3.0	3.0890	0.0729	0.5344	3.0489	0.5404	0.2915
$\lambda_{3,1_2}/\lambda_{2,1_2}$	3.0	3.0890	0.0729	0.5344	3.0489	0.5404	0.2915
$\lambda_{3,1_3}/\lambda_{2,1_3}$	1.667	1.8167	0.0384	0.1685	1.7355	0.3470	0.1240
$\lambda_{3,2_1}/\lambda_{2,1_1}$	3.0	3.0890	0.0729	0.5344	3.0489	0.5404	0.2915
$\lambda_{3,2_2}/\lambda_{2,1_2}$	5.0	5.4796	0.1538	2.5725	5.1762	1.1135	1.2584
$\lambda_{3,3_1}/\lambda_{2,1_1}$	5.0	5.4796	0.1538	2.5725	5.1762	1.1135	1.2584

Baseline hazards $\lambda_{2,1_1} = \lambda_{2,1_2} = 0.10$, $\lambda_{2,1_3} = 0.30$.

Some hazard ratio estimates are identical because both their numerator and their denominators are identical.

[a] The estimates of $\lambda_{2,1_1}$ and $\lambda_{2,2_1}$ are identical.

better. Again, for similar reasons as mentioned for simulation II, some of the hazard ratio estimates are identical in the table.

In summary, the piecewise exponential survival trees not only performed well on three scenarios of the time-dependent covariate but also did reasonably well on a continuously proportional hazards setting. The Cox model with specified baselines and sophisticated codings was efficient, and the proposed method was nearly as efficient as the Cox method. In the simulation studies, the proposed trees method showed some strength. Our example demonstrated that even if a step function for time was added, the proposed trees method was capable of coping with continuous hazards.

4. Stanford Heart Transplant Data

The Stanford heart transplant data is a classic survival data set with the feature that a patient's treatment might change during the study. This well-known data set has been considered and analyzed many times (Mantel and Byar, 1974; Turnbull, Brown, and Hu, 1974; Crowley and Hu, 1977; Kalbfleisch and Prentice, 1980; Aitkin, Laird, and Francis, 1983; Arjas, 1986). Both Crowley and Hu (1977) and Kalbfleisch and Prentice (1980) applied Cox proportional hazards model to the data, using a single model to accommodate both pretransplant and posttransplant survival. Transplantation was defined as a time-dependent covariate. Aitkin et al. (1983) treated pretransplant and posttransplant survival separately and employed both fully parametric modeling (using exponential, Weibull, and lognormal distributions to represent survival or waiting time) and a semiparametric approach using the piecewise exponential distribution. Separate hazard functions were fitted to pretransplant and posttransplant survival using all appropriate covariates, and the effect of transplant was assessed conditional on the time of transplant and the covariate set. The waiting time for transplant was also examined. Obviously, this analysis gives people more details about the association of covariates and underlying hazards. One can not only see whether the transplant affects survival but can also discover changes of the hazards at different time periods.

This content downloaded from
210.117.171.92 on Tue, 27 Jul 2021 04:53:15 UTC
All use subject to https://about.jstor.org/terms

Crowley and Hu (1977) found that age had relatively little effect on survival of the patients before transplantation but that older patients were less able to survive the surgery. Kalbfleisch and Prentice (1980) also found that transplantation was beneficial for younger patients. Their results also indicated that the overall quality of patients being admitted to the study improved with time but that the survival time of the transplanted patients was not improving at the same rate. Aitkin et al. (1983) concluded that younger transplanted patients survived longer, especially those enrolling early in the program.

We applied the piecewise exponential survival trees method to this data set to explore the simultaneous effect of several covariates and whether transplantation is likely to prolong survival. Our analysis illustrates how the piecewise exponential survival trees method works in practice and will lead to a better understanding of the data.

The data are taken from *The Statistical Analysis of Failure Time Data,* by Kalbfleisch and Prentice (1980). This study used 103 patients, 69 of whom received transplants. The censoring rate was 27%. One patient was omitted from our analysis for the same reason given by Aitkin et al. (1983). The response variable ST is survival time. The observed survival time is indicated to be censored or uncensored by the survival status SC (1 = dead, 0 = alive). The time-independent covariates considered in the analysis include the patient's age (AGE), year of acceptance to the study ($YEAR$), and previous open heart surgery ($SURG$) (1 = yes, 0 = no). The time-dependent covariate $TRANS$ is defined by the waiting time for transplanting (WT) in the following ways:

$$TRANS(t)_i = \begin{cases} 0, & \text{if } t < WT_i, \\ 1, & \text{if } t \geq WT_i, \end{cases} \qquad i = 1, 2, \ldots, 102.$$

For a patient who had a heart transplant, we can examine his or her survival experiences at two nonoverlapping time intervals, which here are the time of pretransplant and the time of posttransplant. We employed a minimum node size of 25 and included AGE, $YEAR$, and $SURG$ as time-independent covariates and $TRANS$ as a time-dependent covariate in the procedure.

The piecewise exponential trees first cut the time-dependent covariate $TRANS$ at 0 versus 1, which indicates the time of pretransplant versus the time of posttransplant. It further split posttransplant patients by age ($\leq$48 years old versus 49 or older) and pretransplant patients by acceptance year (before 1970 versus after 1970). Then, it split pretransplant patients who were accepted after 1970 by age ($\leq$46 years old versus 46 or older). The right size tree is a tree with five nodes, which is shown in Figure 3. In the tree diagram, circles and squares represent the nonterminal and the terminal nodes, respectively. The upper-level number inside each node is the number of patients in the node, and the number below is the estimated hazard. From the survival tree, we can conclude that pretransplant patients have the worst prognosis if they were accepted to the study before 1970. Their risk was 10 times greater than that of younger patients after transplant. Age

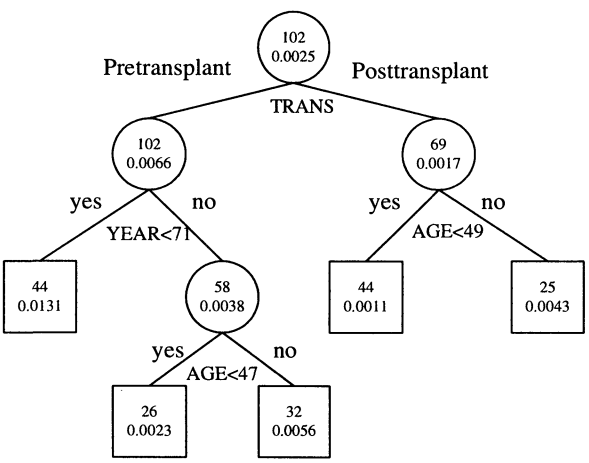

Figure 3. Survival tree of Stanford heart transplant data from the piecewise exponential survival trees method. Circles and squares represent the nonterminal and the terminal nodes, respectively. The upper-level number inside each node is the number of patients in the node, and the number below is the estimated hazard.

This content downloaded from
210.117.171.92 on Tue, 27 Jul 2021 04:53:15 UTC
All use subject to https://about.jstor.org/terms

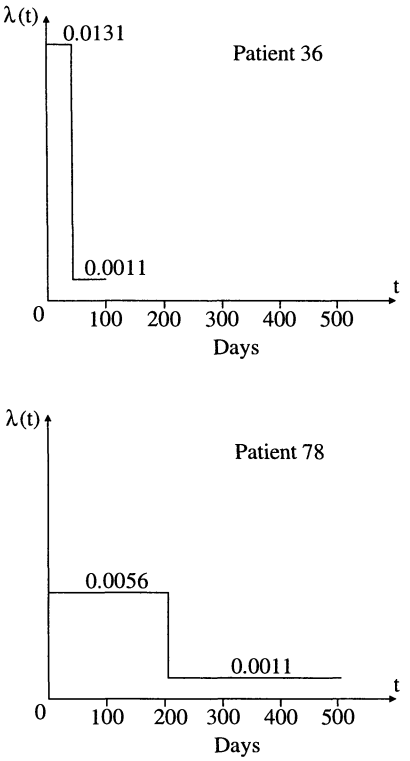

Figure 4. Estimated hazard functions of two patients (ID = 36 and ID = 78). Patient 36 was accepted at age 48 in 1970 and had heart transplant at day 46 and died at day 100. Patient 78 was accepted at age 48 in 1972 and had heart transplant at day 210 and died at day 515.

had a major effect on survival regardless of transplant status. Older posttransplant patients and older pretransplant patients who were accepted to the study after 1970 had similar risks, which were higher than the risk for younger patients. Younger patients benefitted from the transplant, as it reduced their risk by at least 50%.

From the terminal nodes, we can easily estimate a patient's hazard function. Figure 4 shows estimated hazard functions of two patients (ID = 36 and ID = 78). Patient 36 was accepted at age 48 in 1970 and had a heart transplant at day 46 and died at day 100. Patient 78 was accepted at age 48 in 1972 and had a heart transplant at day 210 and died at day 515. Because the time-dependent covariate was split, these two patients appeared in both the pretransplant node and the posttransplant node. This demonstrates the difference between the proposed model and other tree-based models.

In general, our results are consistent with others. As other authors have done (Crowley and Hu, 1977), we treated the patient population as homogeneous. Because randomization was not employed, the possible bias in the selection of patients for transplantation might be inevitable.

5. Discussion

The purpose of this study was to further incorporate time-dependent covariates into tree-based methods for the analysis of survival data. On the basis of a piecewise exponential structure, a new survival trees method has been proposed to handle time-dependent covariates appropriately. As a more flexible alternative to the work of Davis and Anderson (1989), time is incorporated into the model as an argument of the hazard function, which is treated as a time-dependent covariate in computation. Thus, it can handle the situations all other survival trees methods can handle, and the results are comparable. Even with a step function approximation, including time in the model might give a better model fit. In general, the proposed method is a nonparametric approach.

We emphasize the importance and the necessity of including time-dependent covariates in survival analysis whenever applicable. For example, with time-dependent covariates in medical application, the analysis could lead to a better definition of disease prognosis. Another example is to use

This content downloaded from
210.117.171.92 on Tue, 27 Jul 2021 04:53:15 UTC
All use subject to https://about.jstor.org/terms

time-dependent covariates to model changes from allocated treatment (White and Pocock, 1996). Generally, time-dependent covariates can easily be included in the Cox regression model and in the proposed method. However, with time-dependent covariates in the Cox model, it is usually difficult to estimate the baseline hazard, which results in a difficulty in obtaining the hazard function of an individual subject. In contrast, in the proposed model the hazard function for each subject can be estimated automatically. We believe that the survival trees model that includes time-dependent covariates, as proposed in this paper, is a useful alternative to the Cox model.

Simulations were conducted on a variety of models. The proposed trees method was nearly as efficient as Cox regression under the continuous proportional hazards model and performed well in other situations. It should be noted that dummy variables were created to model correctly for the Cox model throughout the simulations. Because we were aware of the data structures in our simulations, we could construct precisely the dummy variables for the Cox model to capture the underlying hazards. However, this might not always be possible in practice, especially when some continuous covariates exist and a log-linear relationship does not hold. On the other hand, the proposed method does not have such problems.

From our limited experience, a sample size of 400 or more with a moderate censoring rate (20% or less) seems to be adequate for the proposed trees method. When the number of covariates is small, it might work adequately for a smaller sample size, as in the Stanford heart transplant data, whereas a larger sample size might be required if the number of covariates is large. In general, larger sample sizes are needed for data with more time-dependent covariates than for those with fewer time-dependent covariates. Also, as in the other methods of survival analysis, to result in the same accuracy more subjects need to be collected for more heavily censored data.

Acknowledgements

The authors thank Dr Julia Austin, Dr Kathryn Kimmel, Ms Cheryl Stein, and Ms Wendy Conner for reviewing the manuscript and the associate editor and two referees for their valuable comments and suggestions.

Résumé

Les méthodes par arbre sont des alternatives non paramétriques au modèle de régression semi-paramétrique de Cox pour l'analyse de la survie. Dans ce papier, une méthode par arbre est proposée pour les données de survie censurées avec des covariables dépendant du temps. La méthode proposée est entièrement nonparamétrique et suppose un modèle très général pour la fonction de risque instantané. L'algorithme de partition récursive utilise la procédure d'estimation par la vraisemblance sous une structure exponentielle par morceaux pour augmenter les arbres, et qui prend en compte les covariables dépendant du temps d'une façon parallèle aux covariables ne dépendant pas du temps. En général, l'estimation du risque instantané à un noeud donne le risque pour un groupe d'individus pendant une période de temps spécifique. Dans la procédure de sélection par arbre sont considérées à la fois des techniques de validation croisée et de rééchantillonnage par bootstrap. Il est montré par simulation et par application à des données réelles que la méthode proposée a de bonnes performances.

References

Ahn, H. and Loh, W.-Y. (1994). Tree-structured proportional hazards regression modeling. *Biometrics* **50,** 471–485.

Aitkin, M., Laird, N., and Francis, B. (1983). A reanalysis of the Stanford heart transplant data (with comments). *Journal of the American Statistical Association* **78,** 264–294.

Andersen, P. K. (1991). Survival analysis 1982–1991: The second decade of the proportional hazards regression model. *Statistics in Medicine* **10,** 1931–1941.

Arjas, E. (1986). Stanford heart transplantation data revisited: A real-time approach. In *Modern Statistical Underline Methods in Chronic Disease Epidemiology*, S. H. Moolgavkar and R. L. Prentice (eds), 65–81. New York: Wiley.

Arjas, E. (1988). A graphical method for assessing goodness of fit in Cox's proportional hazards model. *Journal of the American Statistical Association* **83,** 204–212.

Bacchetti, P. and Segal, M. R. (1995). Survival trees with time-dependent covariates: Application to estimating changes in the incubation period of AIDS. *Lifetime Data Analysis* **1,** 35–47.

Breiman, L., Friedman, J. H., Olshen, R. A., and Store, C. J. (1984). *Classification and Regression Trees.* Belmont, California: Wadsworth.

This content downloaded from
210.117.171.92 on Tue, 27 Jul 2021 04:53:15 UTC
All use subject to https://about.jstor.org/terms

Collett, D. (1994). *Modelling Survival Data in Medical Research.* London: Chapman and Hall.

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

Cox, D. R. and Oakes, D. (1984). *Analysis of Survival Data.* London: Chapman and Hall.

Crowley, J. and Hu, M. (1977). Covariance analysis of heart transplant survival data. *Journal of the American Statistical Association* **72,** 27–36.

Davis, P. B. and Anderson, J. R. (1989). Exponential survival trees. *Statistics in Medicine* **8,** 947–961.

Efron, B. (1982). The jackknife, the bootstrap and other resampling plans. *SIAM* **38.**

Efron, B. and Tibshirani, R. J. (1993). *An Introduction to the Bootstrap.* New York: Chapman and Hall.

Gordon, L. and Olshen, R. A. (1985). Tree-structured survival analysis. *Cancer Treatment Reports* **69,** 1065–1069.

Harrington, D. P. and Fleming, T. R. (1982). A class of rank test procedures for censored survival data. *Biometrika* **69,** 553–566.

Kalbfleisch, J. D. and Prentice, R. L. (1980). *The Statistical Analysis of Failure Time Data.* New York: Wiley.

LeBlanc, M. and Crowley, J. (1992). Relative risk trees for censored survival data. *Biometrics* **48,** 411–425.

LeBlanc, M. and Crowley, J. (1993). Survival trees by goodness of split. *Journal of the American Statistical Association* **88,** 457–467.

Morgan, J. N. and Sonquist, J. A. (1963). Problems in the analysis of survery data and a proposal. *Journal of the American Statistical Association* **58,** 415–434.

Segal, M. R. (1988). Regression trees for censored data. *Biometrics* **44,** 35–47.

Therneau, T. M., Grambsch, P. M., and Fleming, T. R. (1990). Martingale-based residuals for survival models. *Biometrika* **77,** 147–160.

Turnbull, B., Brown, B. W., and Hu, M. (1974). Survivorship analysis of heart transplant data. *Journal of the American Statistical Association* **69,** 74–80.

White, I. R. and Pocock, S. J. (1996). Statistical reporting of clinical trials with individual changes from allocated treatment. *Statistics in Medicine* **15,** 246–262.

Received December 1994; *revised June* 1997; *accepted November* 1997.

This content downloaded from
210.117.171.92 on Tue, 27 Jul 2021 04:53:15 UTC
All use subject to https://about.jstor.org/terms