



Multivariate exponential survival trees and their application to tooth prognosis

Juanjuan Fan^{a,*}, Martha E. Nunn^b, Xiaogang Su^c

^a Department of Mathematics and Statistics, San Diego State University, San Diego, CA 92182, USA

^b Department of Health Policy and Health Services Research, Goldman School of Dental Medicine, Boston University, Boston, MA 02118, USA

^c Department of Statistics and Actuarial Science, University of Central Florida, Orlando, FL 32816, USA

ARTICLE INFO

Article history:

Received 11 April 2008

Received in revised form 30 August 2008

Accepted 6 October 2008

Available online 1 November 2008

ABSTRACT

This paper is concerned with developing rules for assignment of tooth prognosis based on actual tooth loss in the VA Dental Longitudinal Study. It is also of interest to rank the relative importance of various clinical factors for tooth loss. A multivariate survival tree procedure is proposed. The procedure is built on a parametric exponential frailty model, which leads to greater computational efficiency. We adopted the goodness-of-split pruning algorithm of [LeBlanc, M., Crowley, J., 1993. Survival trees by goodness of split. *Journal of the American Statistical Association* 88, 457–467] to determine the best tree size. In addition, the variable importance method is extended to trees grown by goodness-of-fit using an algorithm similar to the random forest procedure in [Breiman, L., 2001. Random forests. *Machine Learning* 45, 5–32]. Simulation studies for assessing the proposed tree and variable importance methods are presented. To limit the final number of meaningful prognostic groups, an amalgamation algorithm is employed to merge terminal nodes that are homogeneous in tooth survival. The resulting prognosis rules and variable importance rankings seem to offer simple yet clear and insightful interpretations.

© 2008 Elsevier B.V. All rights reserved.

1. Introduction

Accurate assignment of tooth prognosis is important to dental clinicians for several reasons. Firstly, it is necessary for the periodontist to devise an appropriate periodontal treatment plan to eliminate local risk factors and halt the progression of periodontal destruction. Secondly, a general dentist needs an accurate tooth prognosis in order to determine the feasibility for the long-term success of any extensive restorative dental work, such as endodontic therapy (root canal), multiple crowns, or fixed bridges, all of which can be prohibitively expensive, especially if the underlying periodontal outlook for the teeth involved is questionable. Lastly, insurance companies are increasingly relying upon various indicators of tooth prognosis to determine the cost-effectiveness of benefits paid for particular procedures.

The ultimate outcome for judging a tooth's condition at any given time is the probability that the tooth will survive in a state of comfort, health, and function in the foreseeable future. Unfortunately, no universal system for assignment of tooth prognosis currently exists. Hence, the assessment of a tooth's condition can vary widely among dental professionals. Typically, a dental clinician will record and weigh a myriad of commonly taught clinical factors in the process of assigning a prognosis. The relative importance that a practitioner assigns to each factor in assessing a tooth's condition is based upon the clinician's training, past clinical experience, and the technical expertise of that particular dentist in treating any condition or pathology present. The subjective nature of this sort of tooth assessment presents difficulty not only for patients, insurance

* Corresponding author. Tel.: +1 619 594 2601.

E-mail address: jjfan@sciences.sdsu.edu (J. Fan).

companies, and other third-party payment systems in determining the cost-effectiveness of treatment, but it also creates confusion among dentists since no clear, objective, evidence-based standards have ever been established.

Since teeth within a person are naturally correlated, the resultant data involve multivariate survival times, which require multivariate extensions of established statistical methodology to accurately develop criteria for assignment of tooth prognosis. In addition, predictors of tooth loss include tooth-level factors, such as the presence of calculus or a large amalgam restoration; subject-level factors, such as smoking, poor oral hygiene, and diabetes; and factors that can be considered on both a patient level and a tooth level, such as alveolar bone loss or plaque.

Previous studies investigating tooth prognosis have focused on assessing which clinical factors are statistically significant in predicting tooth loss as a result of periodontal disease. For instance, in a longitudinal study of 100 well-maintained periodontal patients, McGuire and Nunn (1996a,b) found that many commonly taught clinical factors are significantly associated with tooth loss from periodontal disease. In addition, periodontal prognosis based on clearly defined, objective criteria were also significant predictors of tooth loss from periodontal disease. However, assigned periodontal prognosis failed to accurately account for a tooth's future condition in many cases.

The tree-based method has a nonparametric flavor, making few statistical assumptions. Among its many other attractive features, the hierarchical tree structure automatically groups observations with meaningful interpretation, which renders it an excellent research tool in medical prognosis or diagnosis. Tree-based methods for univariate (or uncorrelated) survival data have been proposed by many authors. See, for example, Gordon and Olshen (1985), Ciampi et al. (1986), Segal (1988), Butler et al. (1989), Davis and Anderson (1989), and LeBlanc and Crowley (1992, 1993). More recently, multivariate (or correlated) survival tree methods have also been proposed. In particular, Su and Fan (2004) and Gao et al. (2004) proposed procedures that make use of the semi-parametric frailty model developed by Clayton (1978) and Clayton and Cuzick (1985).

The Dental Longitudinal Study (DLS) contains a large database of subjects from various socioeconomic backgrounds and differing access to medical and dental care, who have been followed for over 30 years. These subjects are volunteers in a long-term Veterans Administration (VA) study of aging health, but they are not VA patients and receive their medical and dental care in the private sector. Such an extended time of follow-up provides a unique opportunity to observe a large number of tooth loss events.

In order to apply the recursive partitioning algorithms to such a large database, computational efficiency becomes particularly challenging. In this article, we propose a multivariate exponential survival tree procedure. It splits data using the score test statistic derived from a parametric exponential frailty model, which allows for fast evaluation of splits. We then adopt the goodness-of-split pruning algorithm of LeBlanc and Crowley (1993) to determine the optimal tree size. In addition, variable importance rankings are established using a novel random forest algorithm similar to Breiman (2001). We note that (1) the proposed tree method can be viewed as a multivariate extension of the exponential survival trees for independent failure time data studied by Davis and Anderson (1989), with dependence among teeth from the same subject accounted for by a frailty term; (2) the semi-parametric frailty model based tree method by Su and Fan (2004) generally does not work for tooth loss data because such data typically contain many identical censoring times and the frailty model estimation encounters convergence problems; (3) it is the fast computation provided by the parametric multivariate exponential approach that allows us to run the random forest algorithm on a large data set.

The remainder of the paper is organized as follows. In Section 2, the proposed tree and random forest procedures are presented in detail. Section 3 contains simulation studies designed for assessing the multivariate exponential survival tree method and the related variable importance method. In Section 4, we apply the proposed tree and random forest procedures to the analysis of the DLS data set for developing rules for assignment of tooth prognosis and for establishing variable importance rankings for tooth loss. Section 5 concludes the paper with a brief discussion. The interested reader may contact Xiaogang Su (xiaosu@mail.ucf.edu) or Juanjuan Fan (jjfan@sciences.sdsu.edu) for the R programs developed for this paper.

2. The tree procedure

Suppose that there are n units (i.e., subjects) in the data, with the i th unit containing K_i failure times (i.e., teeth). Let F_{ik} be the underlying failure time for the k th member of the i th unit and let C_{ik} be the corresponding underlying censoring time. The data thus consist of $\{(X_{ik}, \Delta_{ik}, \mathbf{Z}_{ik}) : i = 1, 2, \dots, n; k = 1, 2, \dots, K_i\}$, where $X_{ik} = \min(F_{ik}, C_{ik})$ is the observed follow-up time; $\Delta_{ik} = I(F_{ik} \leq C_{ik})$ is the failure indicator, which is 1 if $F_{ik} \leq C_{ik}$ and 0 otherwise; and $\mathbf{Z}_{ik} \in \mathcal{R}^p$ denotes the p -dimensional covariate vector associated with the k th member of the i th unit. The failure times from the same unit are likely to be correlated. In order to ensure identifiability, we assume that the failure time vector $(F_{i1}, \dots, F_{iK_i})'$ is independent of the censoring time vector $(C_{i1}, \dots, C_{iK_i})'$ conditional on the covariate vector $(\mathbf{Z}'_{i1}, \dots, \mathbf{Z}'_{iK_i})'$ for $i = 1, \dots, n$.

To build a tree model T , we follow the convention of classification and regression trees (CART; Breiman et al., 1984), which consists of three steps: first growing a large initial tree, then pruning it to obtain a nested sequence of subtrees, and finally determining the best tree size via validation.

2.1. Growing a large initial tree

We formulate the hazard function of failure time F_{ik} for the k th member of the i th unit by the following exponential frailty model

$$\lambda_{ik}(t) = \exp \{ \beta_0 + \beta_1 z_{ik} \} \cdot w_i \quad (1)$$

for $k = 1, \dots, K_i$ and $i = 1, \dots, n$, where $\exp(\beta_0)$ is an unknown constant baseline hazard, β_1 is an unknown regression parameter, and $z_{ik} = I(Z_{ikj} \leq c)$ with $I(\cdot)$ being the indicator function. For a constant change point c , $Z_{ikj} \leq c$ induces a binary partition of the data according to a continuous covariate Z_j . If Z_j is categorical, then the form of $Z_j \in A$ is considered, where A can be any subset of its categories. For simplicity and interpretability, we restrict ourselves to splits on a single covariate.

The frailty w_i is a multiplicative random effect term that explicitly models the intra-unit dependence. It corresponds to some common unobserved characteristics shared by all members in the i th unit. It is assumed that, given the frailty w_i , failure times within the i th unit are independent. The frailty term w_i is assumed to follow the gamma distribution with mean one and an unknown variance v . Namely, $w_i \sim \text{gamma}(1/v, 1/v)$ with density function

$$g(w, v) = \frac{w^{(1/v)-1} \exp(-w/v)}{\Gamma(1/v)v^{1/v}}, \quad v \geq 0.$$

To split data, we consider the score test for assessing $H_0 : \beta_1 = 0$. Estimation of the null model (i.e., model (1) under H_0) is outlined in [Appendix A](#). Let l denote the integrated log-likelihood function associated with model (1) and $(\hat{\beta}_0, \hat{v})$ the MLE of (β_0, v) from the null model. Also define $\theta = (\beta_0, v)'$ and $m_i = \sum_{k=1}^{K_i} X_{ik} Z_{ik}$. Then the score test statistic can be derived as

$$S = \frac{U^2}{i_{\beta_1\beta_1} - i_{\beta_1\theta} \mathbf{I}_0^{-1} i_{\beta_1\theta}} \Big|_{(\beta_0, v) = (\hat{\beta}_0, \hat{v})}, \quad (2)$$

where

$$\begin{aligned} U &= \frac{\partial l}{\partial \beta_1} \Big|_{\beta_1=0} = \Delta_{..} - \exp(\beta_0) \sum_{i=1}^n \frac{m_i A_i}{B_i}, \\ i_{\beta_1\beta_1} &= - \frac{\partial^2 l}{\partial \beta_1^2} \Big|_{\beta_1=0} = \exp(\beta_0) \sum_{i=1}^n \frac{m_i A_i B_i - \exp(\beta_0) m_i^2 A_i}{B_i^2}, \\ i_{\beta_1\theta} &= (i_{\beta_1\beta_0}, i_{\beta_1 v})' = \left(- \frac{\partial^2 l}{\partial \beta_1 \partial \beta_0}, - \frac{\partial^2 l}{\partial \beta_1 \partial v} \right)' \Big|_{\beta_1=0}, \\ i_{\beta_1\beta_0} &= \exp(\beta_0) \sum_{i=1}^n \frac{m_i A_i}{B_i^2} \{B_i - \exp(\beta_0) X_i\}, \\ i_{\beta_1 v} &= \frac{\exp(\beta_0)}{v^2} \sum_{i=1}^n \frac{m_i}{B_i^2} (A_i - B_i). \end{aligned}$$

In the above expressions, \mathbf{I}_0 is the Fisher information matrix for the null model (see [Appendix A](#) for definitions of $\Delta_{..}$, A_i , and B_i). Since z_{ik} enters S only through m_i , evaluation of the score test statistic for each split is computationally fast. Note that, for a given split, S follows a $\chi^2(1)$ distribution.

The best split is the one that corresponds to the maximum score test statistic among all permissible splits. According to the best split, the whole data is divided into two child nodes. The same procedure is then applied to either of the two child nodes. Recursively proceeding through these steps results in a large initial tree, denoted by T_0 .

2.2. Split-complexity pruning

The final tree model could be any subtree of T_0 . However, the number of subtrees becomes massive even for a moderately-sized tree. To narrow down our choices, we follow CART's ([Breiman et al., 1984](#)) pruning idea to truncate the “weakest link” of the large initial tree T_0 iteratively. Since we employ the maximum score statistic criterion to grow the tree, which provides a natural goodness-of-split measure of each internal node, this allows us to adopt the split-complexity measure of [LeBlanc and Crowley \(1993\)](#):

$$G_\alpha(T) = G(T) - \alpha \cdot |T - \tilde{T}|, \quad (3)$$

where \tilde{T} denotes the set of all terminal nodes of tree T , $T - \tilde{T}$ is the set of all internal nodes, α is termed as complexity parameter, and $G(T) = \sum_{h \in T - \tilde{T}} S(h)$ sums up the maximum score test statistic at every internal node of tree T . Here, $G(T)$ provides an overall goodness-of-split assessment, penalized by tree complexity, i.e., $|T - \tilde{T}|$, the total number of internal nodes.

As the complexity penalty α increases from 0, there will be a link or internal node h that first becomes ineffective because T_h , the branch whose root node is h , is inferior compared to h as a single terminal node. This link is then termed as the “weakest link” and the branch with the smallest threshold value is truncated first.

We start with the large initial tree T_0 . For any link or internal node h of T_0 , calculate the threshold

$$g(h) = \frac{\sum_{i \in T_h - \tilde{T}_h} S(i)}{|T_h - \tilde{T}_h|},$$

where $S(i)$ represents the maximum score test statistic associated with internal node i . Then the “weakest link” h^* in T_0 is the internal node such that $g(h^*) = \min_{h \in T_0 - \tilde{T}_0} g(h)$. Let T_1 be the subtree after pruning off h^* , namely, $T_1 = T_0 - T_{h^*}$. Also, record $\alpha_1 = g(h^*)$. Then we apply the same procedure to truncate T_1 and obtain the subtree T_2 and α_2 . Continuing to do so leads to a decreasing sequence of subtrees $T_0 \succ T_1 \succ \dots \succ T_M$, where T_M is the null tree containing the root node only and the notation \succ means “subtree”, as well as a sequence of increasing α values, $0 = \alpha_0 < \alpha_1 < \dots < \alpha_M < \infty$. Note that all the quantities $S(i)$ have been computed when growing T_0 , which makes the above pruning procedure computationally very fast.

The optimality of this split-complexity algorithm is established by Breiman et al. (1984) and LeBlanc and Crowley (1993). It can be shown that, for any given $\alpha > 0$ and tree T , there exists a unique smallest subtree of T that maximizes the split-complexity measure. This unique smallest subtree, denoted as $T(\alpha)$, is called the optimally pruned subtree of T for complexity parameter α . It has further been established that the subtree T_m resulting from the above pruning procedure is the optimally pruned subtree of T_0 for any α such that $\alpha_m \leq \alpha < \alpha_{m+1}$. In particular, this is the case for the geometric mean of α_m and α_{m+1} , namely $\alpha'_m = \sqrt{\alpha_m \alpha_{m+1}}$.

2.3. Tree size selection

One best-sized tree needs to be selected from the nested sequence. Again, the split-complexity measure $G_\alpha(T)$ in (3) is the yardstick for comparing these candidates. That is, tree T^* is the best sized if $G_\alpha(T^*) = \max_{m=0, \dots, M} \{G(T_m) - \alpha|T_m - T_m|\}$. LeBlanc and Crowley (1993) suggests that, for selecting the best-sized subtree, α be fixed within the range $2 \leq \alpha \leq 4$, where $\alpha = 2$ is in the spirit of AIC (Akaike, 1974) and $\alpha = 4$ corresponds roughly to the 0.05 significance level for a split under the $\chi^2(1)$ curve. However, due to the adaptive nature of greedy searches, the goodness-of-split measure $G(T_m)$ is over-optimistic and needs to be validated to provide a more honest assessment. Two validation methods have been proposed, the test sample method and the resampling method, depending on the sample size available. In the following, the notation $G(\mathcal{L}_2; \mathcal{L}_1, T)$ denotes the validated goodness-of-split measure for tree T built using sample \mathcal{L}_1 and validated using sample \mathcal{L}_2 .

When the sample size is large, the test sample method can be applied. First divide the whole data into two parts: the learning sample \mathcal{L}_1 and the test sample \mathcal{L}_2 . Then grow and prune the initial tree T_0 using \mathcal{L}_1 . At the stage of tree size determination, the goodness-of-split measure $G(T_m)$ is recalculated or validated as $G(\mathcal{L}_2; \mathcal{L}_1, T_m)$ using the test sample \mathcal{L}_2 . The subtree that maximizes the validated G_α is the best-sized tree.

When the sample size is small or moderate, one may resort to resampling techniques. We adopt a bootstrap method proposed by Efron (1983) for bias correction in the prediction problem (also see LeBlanc and Crowley, 1993). In this method, one first grows and prunes a large initial tree T_0 using the whole data. Next, bootstrap samples \mathcal{L}_b , $b = 1, \dots, B$, are drawn from the entire sample \mathcal{L} . A rough guideline for the number of bootstrap samples is $25 \leq B \leq 100$, following LeBlanc and Crowley (1993). Based on each bootstrap sample \mathcal{L}_b , a tree T_0^b is grown and pruned. Let $T_b(\alpha'_m)$, $m = 1, \dots, M$, denote the optimally pruned subtrees corresponding to the α'_m values, where α'_m , $m = 1, \dots, M$, are geometric means of the α_m values respectively, obtained from pruning T_0 . The bootstrap estimator of $G(T_m)$ is given by

$$G^{(b)}(T_m) = G(\mathcal{L}; \mathcal{L}, T(\alpha'_m)) - \frac{1}{B} \sum_{b=1}^B \{G(\mathcal{L}_b; \mathcal{L}_b, T_b(\alpha'_m)) - G(\mathcal{L}; \mathcal{L}_b, T_b(\alpha'_m))\}. \quad (4)$$

The rationale behind this estimator is as follows. The first part of (4), $G(\mathcal{L}; \mathcal{L}, T(\alpha'_m))$, uses the same sample for both growing the tree and calculating the goodness-of-split measure, so the resulting estimate G is over-optimistic. The second part of (4) aims to correct this bias. This part is an estimate of the difference between G 's when the same (\mathcal{L}_b) vs. different (\mathcal{L}_b and \mathcal{L}) samples are used for growing/pruning the tree and for recalculating G , and it is averaged over the B bootstrap samples.

2.4. Variable importance ranking via random forests

Variable importance ranking is another attractive feature offered by recursive partitioning. It is a useful technique if one is interested in answering questions such as which factors are most important for tooth loss/survival. Note that this question cannot be fully addressed by simply examining the splitting variables shown in one single final tree structure, as an important variable can be completely masked by another correlated one. There are a few methods available for extracting variable importance information. Here, we develop an algorithm analogous to the procedure used in random forests (Breiman, 2001), which is among the newest and most promising developments in this regard.

The multivariate exponential survival tree presented in previous sections searches for the best split by maximizing between-node difference. In contrast to CART trees which minimize within-node impurity at each split, this kind of trees

Table 1

Models used for assessing the performance of the splitting statistic. The covariate Z_{ik} 's are generated from a discrete uniform distribution over values $(1/50, 2/50, \dots, 50/50)$ and c is the true cutoff point.

	Model form	Name
A	$\lambda_{ik}(t) = \exp(2 + 1_{\{Z_{ik} \leq c\}}) \cdot w_i$, with $w_i \sim \Gamma(1, 1)$	Gamma-frailty
B	$\lambda_{ik}(t) = \exp(2 + 1_{\{Z_{ik} \leq c\}} + w_i)$, with $w_i \sim N(0, 1)$	Log-normal frailty
C	$\lambda_{ik}(t) = \exp(2 + 1_{\{Z_{ik} \leq c\}})$ not absolutely continuous	Marginal
D	$\lambda_{ik}(t) = \exp(2 + 1_{\{Z_{ik} \leq c\}})$ absolutely continuous	Marginal
E	$\lambda_{ik}(t) = \lambda_0(t) \exp(2 + 1_{\{Z_{ik} \leq c\}}) \cdot w_i$, with $w_i \sim \Gamma(1, 1)$, $\lambda_0(t) = \alpha \lambda t^{\alpha-1}$, $\alpha = 1.5$, and $\lambda = 0.1$	Non-constant baseline

constructed is generally termed as trees by goodness-of-split. To extend random forests to trees by goodness-by-split, once again we make use of the goodness-of-split measure $G(T)$ for tree T .

Let V_j denote the importance measure of the j th covariate Z_j for $j = 1, \dots, p$. We construct random forests by taking B bootstrap samples \mathcal{L}_b , $b = 1, \dots, B$ and based on each bootstrap sample, growing trees with a greedy search over a subset of randomly selected m covariates at each split. The resultant trees are denoted as T_b , $b = 1, \dots, B$. The random selection of m covariates at each split not only helps improve computational efficiency but also produces de-correlated trees. For each tree T_b , the b th out-of-bag sample (denoted as $\mathcal{L} - \mathcal{L}_b$), which contains all observations that are not in \mathcal{L}_b , is sent down to compute the goodness-of-split measure $G(T_b)$. Next, the values of the j th covariate in $\mathcal{L} - \mathcal{L}_b$ are randomly permuted. The permuted out-of-bag sample is sent down T_b again to recompute $G_j(T_b)$. The relative difference between $G(T_b)$ and $G_j(T_b)$ is recorded. This is done for every covariate. The procedure is repeated for B bootstrap samples. The importance V_j is then the average of the relative differences over all B bootstrap samples. The rationale is that the j th covariate of the permuted sample contains no predictive value. Hence, a small difference between $G(T_b)$ and $G_j(T_b)$ indicates small predictive power of the original j th covariate whereas a large difference suggests that the original j th covariate has large predictive power and hence is important. The complete algorithm is summarized in [Appendix B](#).

3. Simulation studies

This section contains simulation experiments designed to investigate the performance of the splitting statistic as well as the tree and variable importance procedures under a variety of model configurations. To facilitate the comparison with [Su and Fan \(2004\)](#), we have employed similar model settings.

3.1. Performance of the splitting statistic

We first investigate the performance of the splitting statistic. We generated data from the five models given in [Table 1](#) with 4 correlated failure times per unit, i.e., $K_i \equiv 4$ for all i . Each model involves a covariate Z simulated independently from a discrete uniform distribution over $(1/50, \dots, 50/50)$ and a threshold covariate effect at cutoff point c .

Model A is a gamma-frailty model. Model B is a log-normal frailty model. Models C and D are marginal models, in which the correlated failure times marginally have the specified hazard and jointly follow a multivariate exponential distribution ([Marshall and Olkin, 1967](#)). This joint multivariate exponential distribution can be absolutely continuous or not, which distinguishes Models C and D. One is referred to [Su et al. \(2006\)](#) on simulating different types of multivariate failure times. Model E involves a baseline hazard function from the Weibull distribution, i.e., a non-constant baseline hazard. Inclusion of Models B, C, D and E is useful for investigating the performance of the splitting statistic under model misspecification.

Two censoring rates were used, no censoring and 50% censoring. The 50% censoring rate was achieved by generating the underlying censoring times from the same model as the underlying failure times. We also tried out two different cutoff points: 0.5 (the middle point) and 0.3 (a shifted one). Only one sample size $n = 100$ was considered. Each combination was examined for 500 simulation runs and the relative frequencies of the best cutoff points selected by the maximum score test (2) are presented in [Table 2](#).

It can be seen that the maximum score test statistic does very well in selecting the correct cutoff points, even under model misspecifications. The results are comparable to those presented in [Su and Fan \(2004\)](#). It is not surprising that the performance deteriorates a little with heavier censoring. But the location of the true cutoff point does not seem to have much effect.

3.2. Performance of the tree method

Next we investigate the efficiency of the proposed multivariate exponential tree procedure in detecting data structure. Data were simulated from the five different models given in [Table 3](#). Four covariates, Z_1, \dots, Z_4 , were generated from a discrete uniform distribution over $(1/10, \dots, 10/10)$, where Z_1 and Z_3 are unit-specific (mimicking subject-level covariates) and Z_2 and Z_4 are failure-specific (mimicking tooth-level covariates). Note that covariates Z_3 and Z_4 do not have any effect on the failure times but are included in the tree procedure as noise. Models A' through E' correspond to those in the previous section, except that each contains two additive threshold effects on Z_1 and Z_2 . Only one censoring rate, 50%, was used.

Table 2

Results for assessing the splitting statistic: Relative frequencies (in percentages) of the ‘best’ cutoff points identified by the maximum score test statistic in 500 runs. The class intervals, identified by the interval midpoints, are $[0, 0.15]$, $(0.15, 0.25]$, $(0.25, 0.35]$, $(0.35, 0.45]$, $(0.45, 0.55]$, $(0.55, 0.65]$, $(0.65, 0.75]$, $(0.75, 0.85]$, and $(0.85, 1]$, respectively. Also, $n = 100$ and $K = 4$.

Model	Cutoff c	Censoring rate (%)	Relative frequencies of the ‘best’ cutoff points								
			0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
A	0.5	0	0.0	0.0	0.2	4.0	93.6	2.2	0.0	0.0	0.0
		50	0.0	0.6	2.0	13.6	75.6	6.6	0.6	0.8	0.2
	0.3	0	0.2	3.2	95.0	1.2	0.4	0.0	0.0	0.0	0.0
		50	2.8	9.0	80.6	5.0	1.6	0.4	0.2	0.0	0.4
B	0.5	0	0.0	0.0	0.2	2.8	94.4	2.6	0.0	0.0	0.0
		50	0.2	0.4	1.6	9.6	76.4	6.0	0.8	0.4	4.6
	0.3	0	0.6	2.8	95.0	1.6	0.0	0.0	0.0	0.0	0.0
		50	2.6	9.4	72.8	6.2	0.4	1.0	0.2	0.0	7.4
C	0.5	0	0.0	0.0	0.0	0.0	99.8	0.0	0.0	0.0	0.2
		50	0.0	0.0	0.0	0.0	98.2	0.4	0.2	0.0	1.2
	0.3	0	0.0	0.0	99.2	0.0	0.0	0.0	0.0	0.0	0.8
		50	0.6	3.6	91.0	0.6	0.2	0.8	1.6	1.0	0.6
D	0.5	0	0.0	0.0	0.0	0.0	95.8	0.2	0.0	0.0	4.0
		50	0.0	0.0	0.4	3.6	95.6	0.4	0.0	0.0	0.0
	0.3	0	0.0	0.0	98.0	0.2	0.0	0.0	0.0	0.0	1.8
		50	0.0	3.4	95.8	0.8	0.0	0.0	0.0	0.0	0.0
E	0.5	0	0.0	0.2	0.4	2.6	94.8	1.8	0.2	0.0	0.0
		50	0.2	1.0	3.6	14.0	70.2	9.0	1.2	0.6	0.2
	0.3	0	0.0	2.2	95.2	2.4	0.2	0.0	0.0	0.0	0.0
		50	3.6	14.4	71.8	7.0	2.4	0.4	0.2	0.0	0.2

Table 3

Models used for assessing the performance of the tree procedure. The covariate Z_{ik} ’s are generated from a discrete uniform distribution over values $(1/10, 2/10, \dots, 10/10)$.

	Model form
A'	$\lambda_{ik}(t) = \exp(-1 + 2 \times 1_{\{Z_{ik1} \leq 0.5\}} + 1_{\{Z_{ik2} \leq 0.5\}}) \cdot w_i$, with $w_i \sim \Gamma(1, 1)$
B'	$\lambda_{ik}(t) = \exp(-1 + 2 \times 1_{\{Z_{ik1} \leq 0.5\}} + 1_{\{Z_{ik2} \leq 0.5\}} + w_i)$, with $w_i \sim N(0, 1)$
C'	$\lambda_{ik}(t) = \exp(-1 + 2 \times 1_{\{Z_{ik1} \leq 0.5\}} + 1_{\{Z_{ik2} \leq 0.5\}})$, not absolutely continuous
D'	$\lambda_{ik}(t) = \exp(-1 + 2 \times 1_{\{Z_{ik1} \leq 0.5\}} + 1_{\{Z_{ik2} \leq 0.5\}})$, absolutely continuous
E'	$\lambda_{ik}(t) = \lambda_0(t) \exp(-1 + 2 \times 1_{\{Z_{ik1} \leq 0.5\}} + 1_{\{Z_{ik2} \leq 0.5\}}) \cdot w_i$, with $w_i \sim \Gamma(1, 1)$, $\lambda_0(t) = \alpha \lambda t^{\alpha-1}$, $\alpha = 1.5$, and $\lambda = 0.1$

The five models were examined for a sample size of $n = 300$ and $K_i \equiv 4$, where the first 200×4 observations formed the learning sample \mathcal{L}_1 and the remaining 100×4 observations formed the test sample \mathcal{L}_2 . Three different complexity parameters ($\alpha = 2, 3$, and 4) were used. For each different model and complexity parameter combination, a total of 200 simulation runs were carried out.

Table 4 presents the final tree size (i.e., the number of terminal nodes) selected by the multivariate exponential survival tree procedure. It can be seen that the proposed procedure does fairly well in detecting the true data structure, even in the misspecified cases (models B' – E'). Also note that $\alpha = 4$ yields the best performance under models A' – D' while values of $\alpha = 2$ and $\alpha = 3$ do not seem to impose enough penalty for model complexity. In model E' , $\alpha = 3$ performs slightly better than $\alpha = 4$ and both perform better than $\alpha = 2$. We will use $\alpha = 4$ when analyzing the DLS data.

3.3. Performance of the variable importance procedure

We also evaluated the performance of the variable importance procedure under the same five models A' – E' as presented in Table 3. A sample size of $n = 500$ and $K_i \equiv 4$ was used to generate data from each model. The same 50% censoring rate was applied. The algorithm in Appendix B was run with $B = 200$ bootstrap samples and $m = 2$ randomly selected inputs at each split. Note that only covariates Z_1 and Z_2 are involved in the five models in Table 3 with Z_1 having a stronger effect than Z_2 : the β coefficients are respectively 2 and 1 for the threshold effects based on Z_1 and Z_2 . The covariates Z_3 and Z_4 are not involved in these models. Hence, the true variable importance ranks (from most important to least important) are Z_1 , followed by Z_2 , and followed by Z_3 and Z_4 (equally unimportant). The simulation results presented in Fig. 1 reflect the true variable importance ranks with Z_1 having the largest variable importance measure, Z_2 having the second largest variable importance measure, and Z_3 and Z_4 having near zero or negative importance measures. Near zero or negative importance measures signify unimportant variables in our procedure.

Table 4

Results for assessing the tree procedure: Relative frequencies (in percentages) of the final tree size identified by the multivariate exponential tree procedure in 200 runs. One set of sample sizes was used, 200×4 observations for the learning sample and 100×4 observations for the test sample with $K = 4$ and 50% censoring rate. Three different complexity parameters (α) were compared: 2, 3, and 4.

Model	Complexity α	Final tree size (in Percentage)						
		1	2	3	4	5	6	≥ 7
A'	2	0.0	0.0	0.0	60.0	14.0	7.0	17.0
	3	0.0	0.0	1.0	76.5	10.0	5.0	7.5
	4	0.0	0.0	2.5	87.5	5.5	2.5	2.0
B'	2	0.0	0.5	2.0	63.5	18.0	5.0	11.0
	3	0.0	0.5	3.5	76.0	13.0	4.5	2.5
	4	0.0	0.5	4.5	81.0	9.5	3.0	1.5
C'	2	0.0	2.0	3.0	39.0	6.0	9.5	41.5
	3	0.0	2.0	3.0	51.0	9.0	9.0	26.0
	4	0.0	2.0	3.0	62.5	7.5	8.0	17.0
D'	2	0.0	3.0	1.5	58.5	10.5	12.5	14.0
	3	0.0	3.0	1.5	68.5	9.0	10.0	8.0
	4	0.0	3.0	2.5	70.5	8.0	10.5	5.5
E'	2	0.0	0.0	8.0	59.5	14.5	8.5	8.5
	3	0.0	0.3	12.5	68.0	8.5	6.0	2.0
	4	0.0	6.0	18.0	65.5	6.0	3.0	1.5

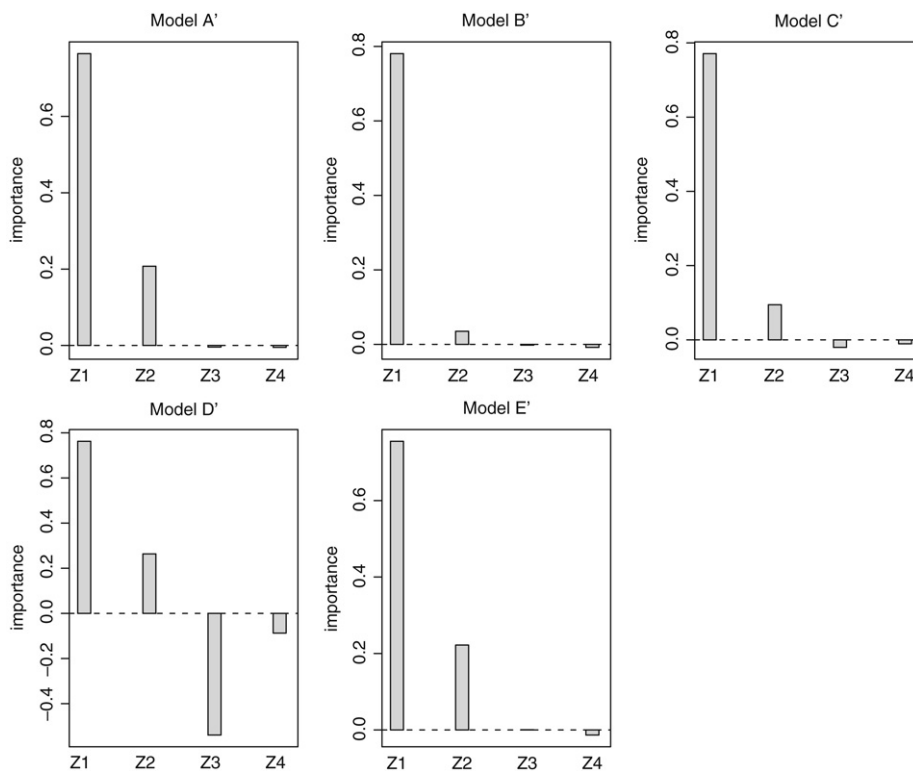


Fig. 1. Variable importance measures based on a data set of 500×4 observations generated from each of Models A' – E': $B = 200$ bootstrap samples were used in random forests and m was set to 2.

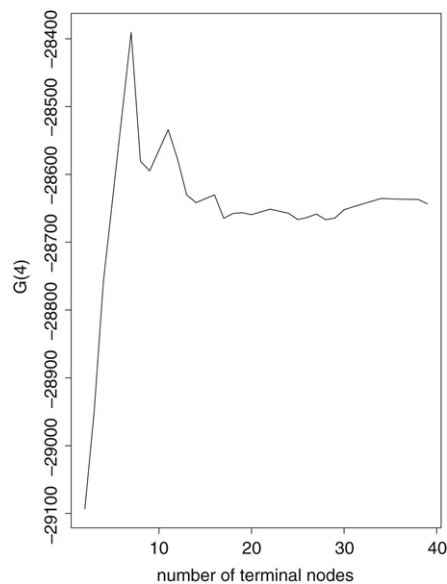
4. Data example

We now apply the proposed tree method to develop prognostic classification rules using the VA Dental Longitudinal Study (DLS) data. We have compiled a subset of 6,463 teeth (from 355 subjects) that have complete baseline information. Tooth loss was observed for 1,103 (17.07%) teeth during the 30 year follow-up. Table 5 presents a description of 19 prognostic factors that we use in the tree procedure.

Table 5

Variable description for the baseline measures in the VA Dental Longitudinal Study (DLS).

Variable	Name	Description
V_1	loc	Tooth type
V_2	nteeth	Total number of teeth present at baseline
V_3	age	Patient age at baseline
V_4	smoke	Smoking status
V_5	education	Educational level
V_6	ses	Socioeconomic level at NAS cycle
V_7	endonum	Number of endodontic teeth
V_8	bonemm	Average bone height (mm)
V_9	abl	Average alveolar bone loss (ABL) score
V_{10}	plaque	Plaque score
V_{11}	mobile	Mobility score
V_{12}	calc	Calculus score, max of above/below gum
V_{13}	ging	Gingival score
V_{14}	maxabl	Maximum tooth-level ABL score
V_{15}	pocket	Probing pocket depth score, max per tooth
V_{16}	dmfscore	DMF score per tooth at baseline
V_{17}	decay	Maximum decay at baseline
V_{18}	endotrtrt	Endodontic treatment
V_{19}	ht	Bone height (%)

**Fig. 2.** Tree size selection for the DLS data: Plot of $G(4)$ versus the tree size.

Since the proposed tree method is computationally fast, the bootstrap method was used for selecting the best-sized tree. A large tree was grown using the whole sample with the minimum node size set at 50 teeth. This tree had 39 terminal nodes. Using the pruning algorithm described earlier, a set of 28 optimally pruned subtrees were identified. With tooth as the sampling unit, 30 bootstrap samples were drawn from the data and, for $m = 1, \dots, 28$, the goodness-of-split measure $G(T_m)$ was evaluated using Eq. (4). To obtain a parsimonious tree model, $\alpha = 4$ was used as the complexity penalty and the $G_4(T_m)$ values for the subtree sequence are presented in Fig. 2. The plot of split-complexity statistics has one peak and the tree with 7 terminal nodes has the largest $G_4(T_m)$.

The final tree is plotted in Fig. 3, where boxes indicate internal nodes and ovals indicate terminal nodes. Inside each box is the splitting rule for the internal node. Observations satisfying this rule go to the left child node and observations not satisfying this rule go to the right child node. The node sizes of internal nodes are omitted because they can be easily calculated once the terminal node sizes are known. For each terminal node, the node size (i.e., the number of teeth) is given within the oval, and the proportion of teeth that are lost is given on the left of each oval.

The final tree in Fig. 3 divides all the teeth into seven groups. However, some groups have similar tooth loss rates. An amalgamation procedure similar to Ciampi et al. (1986) was run to bring down the final number of prognostic groups. To do this, the whole sample was sent down the best-sized tree in Fig. 3 and the robust logrank statistics (Lin, 1994) between every pair of the seven groups were computed. The two terminal nodes giving the smallest robust logrank statistic were merged. This procedure was repeated until the smallest robust logrank statistic was larger than the Bonferroni-adjusted

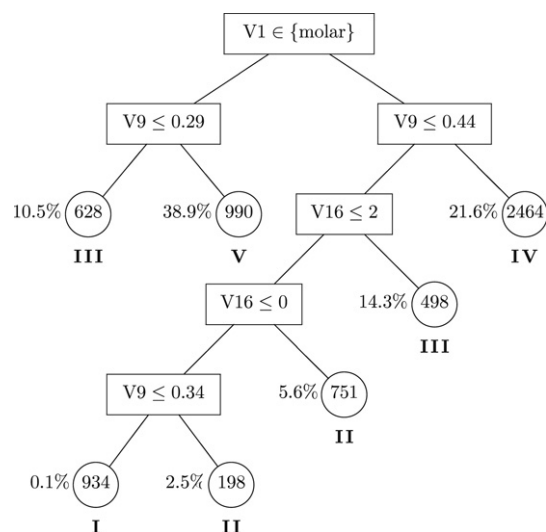


Fig. 3. The best-sized tree for the DLS data. For each internal node denoted by a box, the splitting rule is given inside the box. For each terminal node denoted by an oval, the node size (i.e., number of teeth) is given inside the oval, and the percentage of lost teeth on the left. Underneath terminal nodes in Roman numerals are the five assigned prognostic groups: I—good, II—fair, III—poor, IV—questionable, and V—hopeless.

Table 6

Summary measures and pairwise robust logrank statistics for the five prognostic groups.

Group	Definition	Teeth	Losses (%)		Robust logrank statistic			
					II	III	IV	V
I	Good	934	1	(0.11%)	18.48	38.72	71.03	118.82
II	Fair	949	47	(4.95%)	.	20.45	45.41	103.67
III	Poor	1126	137	(12.17%)	.	.	14.06	66.38
IV	Questionable	2464	533	(21.63%)	.	.	.	42.58
V	Hopeless	990	385	(38.89%)

critical value from a $\chi^2(1)$ distribution. When the algorithm stopped, the number of groups reduced down to five. The amalgamation results are also included in Fig. 3, with Roman numeral I indicating a prognosis of “good”, II indicating a prognosis of “fair”, etc. Summary statistics of the five prognostic groups and the pairwise robust logrank statistics among them are presented in Table 6.

Although nineteen putative prognostic factors were included in the multivariate exponential survival tree procedure, only three of them appeared in the final tree presented in Fig. 3: (1) tooth type: molar vs. non-molar, (2) alveolar bone loss, and (3) tooth-level DMF score: the number of Decayed, Missing, and Filled surfaces for each tooth.

It is particularly interesting to note that the first split was on tooth type with molars having a worse overall outlook than non-molars. This is similar to the results obtained in the studies conducted by McGuire and Nunn (1996a,b), who noted that periodontal prognosis was far more difficult to predict in molars than in non-molars with molars generally having worse overall prognosis than non-molars. One reason for molars having a worse outlook is the potential for the development of furcation involvement. Furcation involvement refers to the situation where alveolar bone loss has progressed to the point on the molar tooth where the roots bifurcate. When furcation involvement develops in a molar, the patient cannot provide adequate oral hygiene to the area which leads to an increased risk of further bone loss as well as an increased risk for dental caries. This sort of furcation involvement only occurs in molars which may explain the model’s distinction between molars and non-molars. In addition, according to the study by McGuire and Nunn (1996b), furcation involvement was one of the most significant factors for predicting tooth loss as a result of periodontal disease. Unfortunately, no measures of furcation involvement were available from the DLS so the only related measure available for the current analysis was tooth type. Hence, it is likely that tooth type in the current study is a proxy measure for furcation involvement, which has long been established to be a significant local etiological risk factor for tooth loss.

The next splits for both molars and non-molars were for average alveolar bone loss. In this case, average alveolar bone loss refers to the subject-level average Schei score. Schei scores range from 0 for “no alveolar bone loss” to 5 for “>80% alveolar bone loss” with values assigned to the mesial and distal areas of each tooth by a trained, calibrated examiner on the basis of examination of the subject’s radiographs. The minimum, median, and maximum of the average ABL score in these data are 0, 0.47, and 2.50, respectively. Three other measures of bone loss that were considered for the model included maximum tooth-level alveolar bone loss, subject-level average bone height, and tooth-level bone height. The maximum tooth-level alveolar bone loss was the maximum Schei score for each tooth. In other words, the maximum Schei score is the maximum of the mesial and distal Schei scores for each tooth. In contrast, the subject-level average bone height and

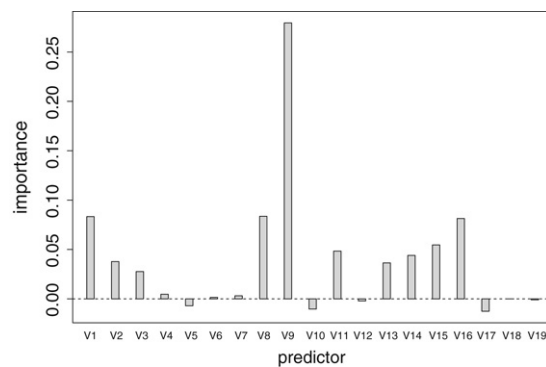


Fig. 4. Variable importance measures in the DLS data obtained via random forests.

the tooth-level bone height were based on digitized radiography. It is not surprising that alveolar bone loss is a significant predictive factor for tooth loss. However, it is surprising that the “best” measure of alveolar bone loss for predicting tooth loss was the average Schei score since a Schei score has always been considered to be less accurate and more subjective than digitized radiography. In addition, since the subject-level average Schei score was selected over the maximum tooth-level Schei score, it would appear that overall bone loss across one’s mouth is a far stronger predictor of tooth loss than an isolated area of vertical bone loss.

The next two splits for non-molars were for tooth-level DMF score. DMF score is actually the number of Decayed, Missing, Filled surfaces for each tooth with values ranging from 0 for a completely sound tooth to 5 for a posterior tooth with all 5 surfaces exhibiting caries experience. In the first split, non-molars with 3 or more surfaces with a positive DMF were assigned a prognosis of “Poor” while non-molars with 1 or 2 surfaces with a positive DMF were assigned a prognosis of “Fair”. Completely sound non-molars (i.e., DMF = 0) were split again according to average alveolar bone loss with sound non-molars with less average alveolar bone loss assigned a prognosis of “Good” while sound non-molars with more average alveolar bone loss were assigned a prognosis of “Fair”.

Finally, Fig. 4 plots the variable importance scores from $B = 200$ bootstrap samples. According to Breiman (2001), variable importance results are “insensitive” to the number of covariates selected to split each node. A choice of $m = 3$ was used in our algorithm. It can be seen that V_9 (abl) stands out as the most important, followed by V_8 (bonemm), V_1 (loc), and V_{16} (dmf score). This matches well with the final tree structure presented in Fig. 3, except that V_8 (bonemm) has been masked out in Fig. 3. Variables V_9 (abl) and V_8 (bonemm) are both measures of average alveolar bone loss across the entire mouth and are essentially different methods for assessing alveolar bone loss with V_9 (abl) being the average Schei score that is assessed by a clinician examining a radiograph and V_8 (bonemm) being the average alveolar bone height from using digitized radiography. Hence, because of the high correlation between V_9 (abl) and V_8 (bonemm), it is expected that only one of these two variables to appear in the final “best tree” model.

From a clinical standpoint, three of the four measures that rank as the most important variables in predicting tooth loss are good indicators of the breakdown of the tooth, as represented by V_{16} (dmf score), and breakdown of the periodontium, as represented by V_9 (abl) and V_8 (bonemm). Since it is well-known that the two leading causes of tooth loss are dental caries and periodontal disease, the statistical importance agrees quite well with widely-held beliefs about clinical importance. The importance of V_1 (loc) reinforces the findings from studies by McGuire and Nunn that show molars are quite different from non-molars when predicting tooth loss. Nevertheless, the method based on random forests provides a more comprehensive evaluation of variable importance by automatically taking into account both the linear and nonlinear effect of a variable and its interaction with others.

It is interesting to note that the two most important variables are both measures of subject-level average alveolar bone loss while the two variables for tooth-level alveolar bone loss, V_{14} (maxabl) and V_{19} (ht), are much less important. From a clinical perspective, the fact that subject-level average alveolar bone loss is substantially more important in predicting tooth loss than tooth-level alveolar bone loss demonstrates that generalized periodontal disease is a stronger predictor of tooth loss than isolated areas of periodontal involvement. The greater importance of subject-level average alveolar bone loss compared to tooth-level alveolar bone loss is also an indication that horizontal bone loss is a more important etiological factor in tooth loss than vertical bone loss since subject-level average alveolar bone loss is a reflection of horizontal bone loss while tooth-level alveolar bone loss is a reflection of vertical bone loss.

5. Discussion

In this paper, we have proposed a multivariate survival tree method. It is an extension of the exponential survival trees by Davis and Anderson (1989) for univariate survival data, but uses a frailty term to adjust for the correlation among survival times from the same cluster. The adoption of the exponential baseline hazard makes it possible to derive a closed-form score test statistic for fast evaluation of tree performance. This paper builds on our work in Su and Fan (2004), but the

computational speed gained from the proposed approach is essential for a frailty model based tree method to be applied to large data sets such as the DLS data because the estimation of a general frailty model is slow and a tree algorithm is feasible only if each model fitting is reasonably fast. Not to mention that it is the computational efficiency achieved by the proposed method that affords the implementation of the random forest algorithm. Though we assume a gamma-frailty term and a constant baseline hazard, simulation studies show that the proposed method works reasonably well under departure from these assumptions.

The application of the proposed tree method to the DLS data yielded five prognostic groups. Three factors were involved in the final tree structure: tooth type (molar vs. non-molar), subject-level average alveolar bone loss score, and tooth-level DMF score. Molars were found to have worse overall prognosis than non-molars. Greater alveolar bone loss was associated with an increased risk of tooth loss. Similarly, the more broken down a tooth is (i.e., the greater the DMF score is), the more likely that the tooth will be damaged further and thus, eventually lost. The application of the variable importance algorithm to the DLS data ranked the subject-level average alveolar bone loss score as the most important variable in predicting tooth loss, followed by the average alveolar bone height, tooth type (molar vs. non-molar), and tooth-level DMF score. The variable of average bone height is another measure of average alveolar bone loss across the entire mouth and hence is masked out in the final tree structure. From a clinical standpoint, the criteria for assignment of prognosis and the variable importance rankings make sense.

Acknowledgement

This research was supported in part by NIH grant R03-DE016924.

Appendix A. Estimation of the null model

Under $H_0 : \beta_1 = 0$, the null frailty exponential model becomes

$$\lambda_{ik}(t) = \exp(\beta_0) \cdot w_i.$$

We first define

$$A_i = 1/v + \Delta_i \quad \text{and} \quad B_i = 1/v + X_i \cdot \exp(\beta_0),$$

where $\Delta_i = \sum_{k=1}^{K_i} \Delta_{ik}$, $X_i = \sum_{k=1}^{K_i} X_{ik}$, and $\Delta_{..} = \sum_{i=1}^n \Delta_i$.

Due to the parametric form of the model, the involved parameters (β_0, v) can be estimated by maximizing the following integrated likelihood

$$l_0 = \prod_{i=1}^n \int_0^\infty \left[g(w_i, v) \prod_{k=1}^{K_i} \{\lambda_{ik}(X_{ik})\}^{\Delta_{ik}} \exp \left\{ - \int_0^{X_{ik}} \lambda_{ik}(t) dt \right\} \right] dw_i.$$

After integrating w_i out, the resultant log-likelihood function for the null model can be written as

$$l_0 = \beta_0 \Delta_{..} + \sum_{i=1}^n \ln \Gamma(A_i) - n \ln \Gamma\left(\frac{1}{v}\right) - \frac{n}{v} \ln v - \sum_{i=1}^n A_i \ln B_i. \quad (5)$$

Optimization of the log-likelihood is done using standard Newton–Raphson procedures. Its gradient components are given by

$$\begin{aligned} \frac{\partial l_0}{\partial \beta_0} &= \Delta_{..} - \sum_{i=1}^n \frac{A_i}{B_i} X_i \cdot \exp(\beta_0) \\ \frac{\partial l_0}{\partial v} &= \frac{1}{v^2} \left\{ - \sum_{i=1}^n \frac{\Gamma'(A_i)}{\Gamma(A_i)} + n \frac{\Gamma'(1/v)}{\Gamma(1/v)} - n + n \ln v + \sum_{i=1}^n \ln B_i + \sum_{i=1}^n \frac{A_i}{B_i} \right\}. \end{aligned}$$

The components in the Fisher information matrix, denoted as \mathbf{I}_0 , are given by

$$\mathbf{I}_0 = \begin{pmatrix} i_{\beta_0 \beta_0} & i_{\beta_0 v} \\ i_{v \beta_0} & i_{vv} \end{pmatrix},$$

where

$$\begin{aligned} i_{\beta_0 \beta_0} &= - \frac{\partial^2 l_0}{\partial \beta_0^2} = \frac{\exp(\beta_0)}{v} \sum_{i=1}^n \frac{A_i}{B_i^2} X_i, \\ i_{\beta_0 v} &= - \frac{\partial^2 l_0}{\partial \beta_0 \partial v} = \frac{\exp(\beta_0)}{v^2} \sum_{i=1}^n \frac{X_i (A_i - B_i)}{B_i^2} = i_{v \beta_0}, \end{aligned}$$

$$i_{vv} = -\frac{\partial^2 l_0}{\partial v^2} = \frac{2n}{v^3} \frac{\Gamma'(1/v)}{\Gamma(1/v)} - \frac{3n}{v^3} + \frac{2n \ln v}{v^3} + \frac{n}{v^4} \frac{\Gamma''(1/v)}{\Gamma'(1/v)} - \frac{n}{v^4} \left\{ \frac{\Gamma'(1/v)}{\Gamma(1/v)} \right\}^2 \\ + \sum_{i=1}^n \left[\frac{2\Gamma'(A_i)}{v^3 \Gamma(A_i)} - \frac{2 \ln B_i}{v^3} - \frac{2A_i}{v^3 B_i} + \frac{1}{v^4} \frac{\Gamma''(A_i)}{\Gamma(A_i)} - \frac{1}{v^4} \left\{ \frac{\Gamma'(A_i)}{\Gamma(A_i)} \right\}^2 - \frac{2}{v^4 B_i} + \frac{A_i}{v^4 B_i^2} \right].$$

Appendix B. Algorithm for computing variable importance via random forests

Initialize all V_j 's to 0.

For $b = 1, 2, \dots, B$, do

- Generate bootstrap sample \mathcal{L}_b and obtain the out-of-bag sample $\mathcal{L} - \mathcal{L}_b$.
- Based on \mathcal{L}_b , grow a tree T_b by searching over m randomly selected inputs at each split and without pruning.
- Send $\mathcal{L} - \mathcal{L}_b$ down T_b to compute $G(T_b)$.
- For all predictors $Z_j, j = 1, \dots, p$, do
 - Permute the values of Z_j in $\mathcal{L} - \mathcal{L}_b$.
 - Send the permuted $\mathcal{L} - \mathcal{L}_b$ down to T_b to compute $G_j(T_b)$.
 - Update $V_j \leftarrow V_j + \frac{G(T_b) - G_j(T_b)}{G(T_b)}$.
- End do.

End do.

Average $V_j \leftarrow V_j/B$.

References

- Akaike, H., 1974. A new look at model identification. *IEEE Transactions on Automatic Control* 19, 716–723.
- Breiman, L., 2001. Random forests. *Machine Learning* 45, 5–32.
- Breiman, L., Friedman, J., Olshen, R., Stone, C., 1984. *Classification and Regression Trees*. Wadsworth International Group, Belmont, CA.
- Butler, J., Gilpin, E., Gordon, L., Olshen, R., 1989. Tree-structured survival analysis, II. Technical Report. Stanford University, Department of Biostatistics.
- Ciampi, A., Thiffault, J., Nakache, J.-P., Asselain, B., 1986. Stratification by stepwise regression, correspondence analysis and recursive partition. *Computational Statistics and Data Analysis* 4, 185–204.
- Clayton, D.G., 1978. A model for association in bivariate life tables and its application in epidemiologic studies of familial tendency in chronic disease incidence. *Biometrika* 65, 141–151.
- Clayton, D.G., Cuzick, J., 1985. Multivariate generalization of the proportional hazards model. *Journal of the Royal Statistical Society, Series A* 148, 82–108.
- Davis, R., Anderson, J., 1989. Exponential survival trees. *Statistics in Medicine* 8, 947–962.
- Efron, B., 1983. Estimating the error rate of a prediction rule: Improvements on cross-validation. *Journal of American Statistical Association* 78, 316–331.
- Gao, F., Manatunga, A.K., Chen, S., 2004. Identification of prognostic factors with multivariate survival data. *Computational Statistics and Data Analysis* 45, 813–824.
- Gordon, L., Olshen, R., 1985. Tree-structured survival analysis. *Cancer Treatment Reports* 69, 1065–1069.
- LeBlanc, M., Crowley, J., 1992. Relative risk trees for censored survival data. *Biometrics* 48, 411–425.
- LeBlanc, M., Crowley, J., 1993. Survival trees by goodness of split. *Journal of the American Statistical Association* 88, 457–467.
- Lin, D.Y., 1994. Cox regression analysis of multivariate failure time data: The marginal approach. *Statistics in Medicine* 13, 2233–2247.
- Marshall, A.W., Olkin, I., 1967. A multivariate exponential distribution. *Journal of American Statistical Association* 62, 30–44.
- McGuire, M.K., Nunn, M.E., 1996a. Prognosis versus actual outcome. II: The effectiveness of commonly taught clinical parameters in developing an accurate prognosis. *Journal of Periodontology* 67, 658–665.
- McGuire, M.K., Nunn, M.E., 1996b. Prognosis versus actual outcome. III: The effectiveness of clinical parameters in accurately predicting tooth survival. *Journal of Periodontology* 67, 666–674.
- Segal, M.R., 1988. Regression trees for censored data. *Biometrics* 44, 35–47.
- Su, X.G., Fan, J.J., 2004. Multivariate survival trees: A maximum likelihood approach based on frailty models. *Biometrics* 60, 93–99.
- Su, X.G., Fan, J.J., Wang, A., Johnson, M., 2006. On simulating multivariate failure times. *International Journal of Applied Mathematics and Statistics* 5, 8–18.